

Attorney's Docket No.: 442-010527-US(PAR)

# 3

PATENT

0360  
10/11/01

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: CAGLAR et al.  
Serial No.: 09/935,119  
Filed: 8/21/01  
For: VIDEO CODING

Group No.:

Examiner:

Commissioner of Patents and Trademarks  
Washington, D.C. 20231

TRANSMITTAL OF CERTIFIED COPY

Attached please find the certified copy of the foreign application from which priority is claimed for this case:

Country : Finland  
Application Number : 20001847  
Filing Date : 21 August 2000

**WARNING:** "When a document that is required by statute to be certified must be filed, a copy, including a photocopy or facsimile transmission of the certification is not acceptable." 37 CFR 1.461 (emphasis added.)

  
SIGNATURE OF ATTORNEY

Clarence A. Green

Reg. No.: 24,622

Type or print name of attorney

Tel. No.: (203) 259-1800

Perman & Green, LLP

Customer No.: 2512

P.O. Address

425 Post Road, Fairfield, CT 06430

NOTE: The claim to priority need be in no special form and may be made by the attorney or agent if the foreign application is referred to in the oath or declaration as required by § 1.63.

CERTIFICATE OF MAILING/TRANSMISSION (37 CFR 1.8a)

I hereby certify that this correspondence is, on the date shown below, being:

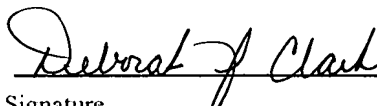
☒ MAILING

deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to the Commissioner of Patents and Trademarks, Washington, D.C. 20231

☐ FACSIMILE

transmitted by facsimile to the Patent and Trademark Office

Date: OCTOBER 12, 2001

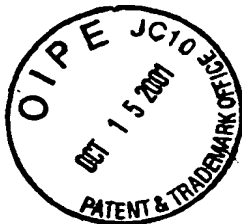
  
Signature

DEBORAH J. CLARK  
(type or print name of person certifying)

(Transmittal of Certified Copy [5-4])

PATENTTI- JA REKISTERIHALLITUS  
NATIONAL BOARD OF PATENTS AND REGISTRATION

Helsinki 25.7.2001



ETUOIKEUSTODISTUS  
PRIORITY DOCUMENT



Hakija  
Applicant

Nokia Mobile Phones Ltd  
Espoo

Patenttihakemus nro  
Patent application no

20001847

Tekemispäivä  
Filing date

21.08.2000

Kansainvälinen luokka  
International class

H04N

Keksinnön nimitys  
Title of invention

"Video coding"  
(Videokoodaus)

Täten todistetaan, että oheiset asiakirjat ovat tarkkoja jäljennöksiä patentti- ja rekisterihallitukselle alkuaan annetuista selityksestä, patenttivaatimuksista, tiivistelmästä ja piirustuksista.

This is to certify that the annexed documents are true copies of the description, claims, abstract and drawings originally filed with the Finnish Patent Office.

  
Pirjo Kaila  
Tutkimussihteeri

Maksu 300,- mk  
Fee 300,- FIM

Osoite: Arkadiankatu 6 A Puhelin: 09 6939 500 Telefax: 09 6939 5328  
P.O.Box 1160 Telephone: + 358 9 6939 500 Telefax: + 358 9 6939 5328  
FIN-00101 Helsinki, FINLAND

## VIDEO CODING

The invention relates to data transmission and is particularly, but not exclusively, related to transmission of data representative of picture sequences, such as video.

- 5 It is particularly suited to transmission over links susceptible to errors and loss of data, such as over the air interface of a cellular telecommunications system.

During the past few years, the amount of multi-media content available through the Internet has increased considerably. Since data delivery rates to mobile terminals  
10 are becoming high enough to enable such terminals to be able to retrieve multi-media content, it is becoming desirable to provide such retrieval from the Internet. An example of a high-speed data delivery system is the General Packet Radio Service (GPRS) of the planned GSM phase 2+.

- 15 The term multi-media as used herein includes both sound and pictures, sound only and pictures only. Sound includes speech and music.

In the Internet, transmission of multi-media content is packet-based. Network traffic through the Internet is based on a transport protocol called the Internet  
20 Protocol (IP). IP is concerned with transporting data packets from one location to another. It facilitates the routing of packets through intermediate gateways, that is, it allows data to be sent to machines that are not directly connected in the same physical network. The unit of data transported by the IP layer is called an IP datagram. The delivery service offered by IP is connectionless, that is IP  
25 datagrams are routed around the Internet independently of each other. Since no resources are permanently committed within the gateways to any particular connection, the gateways may occasionally have to discard datagrams because of lack of buffer space or other resources. Thus, the delivery service offered by IP is a best effort service rather than a guaranteed service.

30 Internet multi-media is typically streamed using the User Datagram Protocol (UDP), the Transmission Control Protocol (TCP) or the Hypertext Transfer Protocol (HTTP). UDP does not check that the datagrams have been received,

does not retransmit missing datagrams, nor does it guarantee that the datagrams are received in the same order as they were transmitted. UDP is connectionless. TCP checks that the datagrams have been received and retransmits missing datagrams. It also guarantees that the datagrams are received in the same order as they were transmitted. TCP is connection orientated.

In order to ensure multi-media content of a sufficient quality is delivered, it can be provided over a reliable network connection, such as TCP, to ensure that received data are error-free and in the correct order. Lost or corrupted protocol data units are retransmitted.

Sometimes re-transmission of lost data is not handled by the transport protocol but rather by some higher-level protocol. Such a protocol can select the most vital lost parts of a multi-media stream and request the re-transmission of those. The most vital parts can be used for prediction of other parts of the stream, for example.

Multi-media content typically includes video. In order to be transmitted efficiently, video is often compressed. Therefore, compression efficiency is an important parameter in video transmission systems. Another important parameter is the tolerance to transmission errors. Improvement in either one of these parameters adversely affects the other and so a video transmission system should have a suitable balance between the two.

Figure 1 shows a video transmission system. The system comprises a source coder which compresses an uncompressed video signal to a desired bit rate thereby producing an encoded and compressed video signal and a source decoder which decodes the encoded and compressed video signal to re-construct the uncompressed video signal. The source coder comprises a waveform coder and an entropy coder. The waveform coder performs lossy video signal compression and the entropy coder losslessly converts the output of the waveform coder into a binary sequence. The binary sequence is conveyed from the source coder to a transport coder which encapsulates the compressed video according to a suitable transport protocol and then transmits it to a receiver comprising a

transport decoder and a source decoder. The data is transmitted by the transport coder to the transport decoder over a transmission channel. The transport coder may also manipulate the compressed video in other ways. For example, it may interleave and modulate the data. After being received by the transport decoder the data is then passed on to the source decoder. The source decoder comprises a waveform decoder and an entropy decoder. The transport decoder and the source decoder perform inverse operations to obtain the re-constructed video signal for display. The receiver may also provide feedback to the transmitter. For example, the receiver may signal the rate of successfully received transmission data units.

A video sequence consists of a series of still images. A video sequence is compressed by reducing its redundant and perceptually irrelevant parts. The redundancy in a video sequence can be categorised as spatial, temporal and spectral redundancy. Spatial redundancy refers to the correlation between neighbouring pixels within the same image. Temporal redundancy refers to the fact that objects appearing in a previous image are likely to appear in a current image. Spectral redundancy refers to the correlation between the different colour components of an image.

Temporal redundancy can be reduced by generating motion compensation data, which describes relative motion between the current image and the previous image (referred to as a reference or anchor picture). Effectively the current image is formed as a prediction from a previous one and the technique by which this is achieved is commonly referred to as motion compensated prediction or motion compensation. In addition to predicting one picture from another, parts or areas of a single picture may be predicted from other parts or areas of that picture.

A sufficient level of compression cannot usually be reached just by reducing the redundancy of a video sequence. Therefore, video encoders also try to reduce the quality of those parts of the video sequence which are subjectively less important. In addition, the redundancy of the encoded bit-stream is reduced by means of

efficient lossless coding of compression parameters and coefficients. The main technique is to use variable length codes.

Video compression methods typically differentiate images on the basis of whether they do or do not utilise temporal redundancy reduction (that is whether they are predicted or not). Compressed images which do not utilise temporal redundancy reduction methods are usually called INTRA or I-frames. INTRA frames are frequently introduced to prevent the effects of packet losses from propagating spatially and temporally. In broadcast situations, INTRA frames enable new receivers to start decoding the stream, that is they provide "access points". Video coding systems typically enable insertion of INTRA frames periodically every  $n$  seconds. It is also advantageous to utilise INTRA frames at natural scene cuts where the image content changes so drastically that temporal prediction from the previous image is unlikely to be successful or desirable in terms of compression efficiency.

Compressed images which do utilise temporal redundancy reduction methods are usually called INTER or P-frames. INTER frames employing motion-compensation are rarely precise enough to allow sufficiently accurate image re-construction and so a spatially compressed prediction error image is also associated with each INTER frame. This represents the difference between the current frame and its prediction.

A group of pictures (GOP) is an INTRA frame and a sequence of temporally predicted pictures predicted from it.

One useful property of coded bit-streams is scalability. In the following, bit-rate scalability is described which refers to the ability of a compressed sequence to be decoded at different data rates. Such a compressed sequence can be streamed over channels with different bandwidths and can be decoded and played back in real-time at different receiving terminals.

Scalable multi-media is typically ordered into hierarchical layers of data. A base layer contains an individual representation of a multi-media clip such as a video sequence and enhancement layers contain refinement data in addition to the base layer. The quality of the multi-media clip progressively improves as enhancement layers are added to the base layer.

Scalability is a desirable property for heterogeneous and error prone environments such as the Internet and wireless channels in cellular communications networks. This property is desirable in order to counter limitations such as constraints on bit rate, display resolution, network throughput and decoder complexity.

In multi-point and broadcast multi-media applications, constraints on network throughput may not be foreseen at the time of encoding. Thus, it is advantageous to encode multi-media content to form a scalable bit-stream. An example of a scalable bit-stream being used in IP multi-casting is shown in Figure 3. Each router (R1-R3) can strip the bit-stream according to its capabilities. In this example, the server has a multi-media clip which can be scaled to at least three bit rates, 120 kbit/s, 60 kbit/s and 28 kbit/s. In the case of a multi-cast transmission (where the same bit-stream is delivered to multiple clients at the same time with as few copies of the bit-stream being generated in the network as possible), it is beneficial from the point of view of network bandwidth to transmit a single bit-rate-scalable bit-stream.

If a sequence is downloaded and played back in different devices each having different processing powers, bit-rate scalability can be used in devices having lower processing power to provide a lower quality representation of the video sequence by decoding only a part of the bit-stream. Devices having higher processing power can decode and play the sequence with full quality. Additionally, bit-rate scalability means that the processing power needed for decoding a lower quality representation of the video sequence is lower than when decoding the full quality sequence. This is a form of computational scalability.

If a video sequence is pre-stored in a streaming server, and the server has to temporarily reduce the bit-rate at which it is being transmitted as a bit-stream, for example in order to avoid congestion in the network, it is advantageous if the server can reduce the bit-rate of the bit-stream whilst still transmitting a useable bit-stream. This is typically achieved using bit-rate scalable coding.

Scalability can be used to improve error resilience in a transport system where layered coding is combined with transport prioritisation. The term transport prioritisation is used to describe mechanisms that provide different qualities of service in transport. These include unequal error protection, which provides different channel error/loss rates, and assigning different priorities to support different delay/loss requirements. For example, the base layer of a scalably encoded bit-stream may be delivered through a transmission channel with a high degree of error protection, whereas the enhancement layers may be transmitted in more error-prone channels.

One problem with scalable multi-media coding is that it often suffers from a worse compression efficiency than non-scalable coding. A high-quality scalable video sequence generally requires more bandwidth than a non-scalable single-layer video sequence of a corresponding quality. However, exceptions to this generality exist, for example temporally scalable B-frames in video compression may improve coding efficiency, especially at high frame rates. B-frames are discussed in the following.

Various video coding standards have been proposed. One such standard, H.263, is an International Telecommunications Union (ITU) video coding recommendation which specifies the bit-stream syntax and the decoding of a bit-stream. Currently, there are two versions of H.263. Version 1 consists of a core algorithm and four optional coding modes. H.263 version 2 is an extension of version 1 which provides twelve negotiable coding modes. H.263 version 3, which is presently under development, is intended to contain two new coding modes and a set of additional supplemental enhancement information code-points.



According to H.263, pictures are coded as luminance and two colour difference (chrominance) components ( $Y$ ,  $C_B$  and  $C_R$ ). The chrominance components are sampled at half resolution along both co-ordinate axes compared to the luminance component.

5

Each coded picture, as well as the corresponding coded bit stream, is arranged in a hierarchical structure with four layers being, from top to bottom, a picture layer, a picture segment layer, a macroblock (MB) layer and a block layer. The picture segment layer can be either a group of blocks layer or a slice layer.

10

The picture layer data contains parameters affecting the whole picture area and the decoding of the picture data. The coded data is arranged in a so-called picture header.

15 By default, each picture is divided into groups of blocks. A group of blocks (GOB) typically comprises 16 subsequential pixel lines. Data for each GOB consist of an optional GOB header followed by data for MBs.

20 If an optional slice structured mode is used, each picture is divided into slices instead of GOBs. Data for each slice consists of a slice header followed by data for MBs.

25 A slice defines a region within a coded picture. Typically, the region is a number of MBs in normal scanning order. There are no prediction dependencies across slice boundaries within the same coded picture. However, temporal prediction can generally cross slice boundaries unless H.263 Annex R (Independent Segment Decoding) is used. Slices can be decoded independently from the rest of the image data (except for the picture header). Consequently, slices improve error resilience in packet-lossy networks.

30

Picture, GOB and slice headers begin with a synchronisation code. No other code word or valid combination of code words can form the same bit pattern as the synchronisation codes. Thus, the synchronisation codes can be used for bit-

stream error detection and re-synchronisation after bit errors. The more synchronisation codes that are added to the bit stream the more error-robust coding becomes.

- 5 Each GOB or slice is divided into MBs. An MB relates to  $16 \times 16$  pixels of luminance data and the spatially corresponding  $8 \times 8$  pixels of chrominance data. In other words, an MB consists of four  $8 \times 8$  luminance blocks and the two spatially corresponding  $8 \times 8$  chrominance blocks.
- 10 A block relates to  $8 \times 8$  pixels of luminance or chrominance data. Block layer data consist of uniformly quantised discrete cosine transform coefficients, which are scanned in zig-zag order, processed with a run-length encoder and coded with variable length codes, as explained in detail in ITU-T recommendation H.263.
- 15 Many video compression schemes also introduce temporally bi-directionally-predicted frames, which are commonly referred to as B-pictures or B-frames. B-frames are inserted between anchor frame pairs and are predicted from either one or both of the anchor frames, as is shown in Figure 2. B-frames are not themselves used as anchor frames, that is other frames are never predicted from
- 20 them. B-frames are used to enhance perceived image quality by increasing the picture display rate. They can be dropped without affecting the decoding of subsequent frames, thus enabling a video sequence to be decoded at different rates according to bandwidth constraints of the transmission network, or different decoder capabilities. Thus, B-frames also provide temporal scalability. Whilst B-
- 25 frames may improve compression performance compared to P-frames, their use requires greater computational and increased memory as well as introducing additional delays.

- 30 Spatial scalability is closely related to another form of scalability referred to as Signal-to-Noise Ratio (SNR) scalability. An example of SNR scalable pictures is shown in Figure 4. SNR scalability involves the creation of multi-rate bit streams. It allows for the recovery of coding errors, or differences, between an original picture and its re-construction. This is achieved by using a finer quantiser to encode the

difference picture in an enhancement layer. This additional information increases the SNR of the overall reproduced picture.

Spatial scalability allows for the creation of multi-resolution bit-streams to meet varying display requirements/constraints. A spatial scalable structure is shown in Figure 5. It is similar to that used in SNR scalability. In spatial scalability, a spatial enhancement layer is used to recover the coding loss between an up-sampled version of the re-constructed layer used as a reference by the enhancement layer, that is the reference layer, and a higher resolution version of the original picture.

For example, if the reference layer has a Quarter Common Intermediate Format (QCIF) resolution, 176x144 pixels, and the enhancement layer has a Common Intermediate Format (CIF) resolution, 352x288 pixels, the reference layer picture must be scaled accordingly such that the enhancement layer picture can be appropriately predicted from it. According to H.263 the resolution is increased by a factor of two in the vertical direction only, horizontal direction only, or both the vertical and horizontal directions for a single enhancement layer. There can be multiple enhancement layers, each increasing picture resolution over that of the previous layer. Interpolation filters used to up-sample the reference layer picture are explicitly defined in H.263. Apart from the up-sampling process from the reference to the enhancement layer, the processing and syntax of a spatially scaled picture are identical to those of an SNR scaled picture. Spatial scalability provides increased spatial resolution over SNR scalability.

In either SNR or spatial scalability, the enhancement layer pictures are referred to as EI- or EP-pictures. If the enhancement layer picture is upwardly predicted from an INTRA picture in the reference layer, then the enhancement layer picture is referred to as an Enhancement-I (EI) picture. In some cases, when reference layer pictures are poorly predicted, over-coding of static parts of the picture can occur in the enhancement layer, requiring an excessive bit rate. To avoid this problem, forward prediction is permitted in the enhancement layer. A picture that is forwardly predicted from a previous enhancement layer picture or upwardly predicted from a predicted picture in the reference layer is referred to as an Enhancement-P (EP) picture. Computing the average of both upwardly and

forwardly predicted pictures can provide a bi-directional prediction option for EP-pictures. Upward prediction of EI- and EP-pictures from a reference layer picture implies that no motion vectors are required. In the case of forward prediction for EP-pictures, motion vectors are required.

5

The scalability mode (Annex O) of H.263 specifies syntax to support temporal, SNR, and spatial scalability capabilities.

10

One problem with conventional SNR scalability coding is termed drifting. Drifting refers to the impact of a transmission error. A visual artefact caused by an error drifts temporally from the picture in the error occurs. Due to motion compensation, the area of the visual artefact may increase from picture to picture. In the case of scalable coding, the visual artefact also drifts from lower enhancement layers to higher layers. The effect of drifting can be explained with reference to Figure 7

15

which shows conventional prediction relationships used in scalable coding. Once an error or packet loss has occurred in an enhancement layer, it propagates to the end of a group of pictures (GOP) since the pictures are predicted for each other in sequence. In addition, since the enhancement layers are based on the base layer, an error in the base layer causes errors in the enhancement layers which are also

20

based on each other. Therefore, this causes a serious drifting problem in higher layers in the prediction frames followed. Even though there may subsequently be sufficient bandwidth to send data to correct an error, the decoder is not able to eliminate the error until the prediction chain is stopped by another INTRA picture starting a new GOP.

25

To deal with this problem, a form of scalability referred to as Fine Granularity Scalability (FGS) has been developed. In FGS a low-quality base layer is coded using a hybrid predictive loop and an (additional) enhancement layer delivers the progressively encoded residue between the re-constructed base layer and the original frame. FGS has been proposed, for example, in MPEG-4 visual standardisation.

30

An example of prediction relationships in fine granularity scalable coding is shown in Figure 6. In a fine granularity scalable video coding scheme, the base-layer video is transmitted in a well-controlled channel to minimise error or packet-loss, in such a way that the base layer is encoded to fit into the minimum channel bandwidth. This minimum is the lowest bandwidth that may occur or may be encountered during operation. All enhancement layers in the prediction frames are coded based on the base layer in the reference frames. Thus, errors in the enhancement layer of one frame do not cause a drifting problem in the enhancement layers of subsequently predicted frames and the coding scheme can adapt to channel conditions. However, since prediction is always based on a low quality base-layer, the coding efficiency of FGS coding is not as good as, and is sometimes much worse than, conventional SNR scalability schemes such as that in H.263 Annex O.

In order to combine the advantages of both FGS coding and conventional layered scalability coding, a hybrid coding scheme shown in Figure 8 has been proposed which is called Progressive FGS (PFGS). There are two points to note. Firstly, in PFGS as many predictions as possible from the same layer are used to maintain coding efficiency. Secondly, a prediction path always uses prediction from a lower layer in the reference frame to enable error recovery and channel adaptation. The first point makes sure that, for a given video layer, motion prediction is as accurate as possible, thus maintaining coding efficiency. The second point makes sure that there is no drifting problem in case of channel congestion, packet loss or packet error. Using this coding structure, there is no need to re-transmit lost/erroneous packets since enhancement layers can be gradually and automatically re-constructed over a period of a few frames.

In Figure 8, frame 2 is predicted from the even layers of frame 1 (that is the base layer and the 2nd layer). Frame 3 is predicted from the odd layers of frame 2 (that is the 1st and the 3rd layer). In turn, frame 4 is predicted from the even layers of frame 3. This odd/even prediction pattern continues. The term group depth is used to describe the number of layers that refer back to a common reference layer. Figure 8 exemplifies a case where the group depth is 2. The group depth can be

changed. If the depth is 1, the situation is essentially equivalent to a traditional scalability scheme. If the depth is equal to the total number of layers, the scheme becomes equivalent to FGS. Thus, the progressive FGS coding scheme illustrated in Figure 8 offers a compromise that provides the advantages of both the previous techniques, such as high coding efficiency and error recovery.

PFGS provides advantages when applied to video transmission over the Internet or over wireless channels. The encoded bit-stream can adapt to the available bandwidth of a channel without significant drifting occurring. Figure 9 shows an example of the bandwidth adaptation property provided by progressive fine granularity scalability in a situation where a video sequence is represented by frames having a base layer and 3 enhancement layers. The thick dot-dashed line traces the video layers actually transmitted. At frame 2, there is significant reduction in bandwidth. The transmitter (server) reacts to this by dropping the bits representing the higher enhancement layers (layers 2 and 3). After frame 2, the bandwidth increases to some extent and the transmitter is able to transmit the additional bits representing two of the enhancement layers. By the time frame 4 is transmitted, the available bandwidth has further increased, providing sufficient capacity for the transmission of the base layer and all enhancement layers again. These operations do not require any re-encoding and re-transmission of the video bit-stream. All layers of each frame of the video sequence are efficiently coded and embedded in a single bit-stream.

The prior art scalable encoding techniques described above are based on a single interpretation of the encoded bit-stream. In other words, the decoder interprets the encoded bit-stream only once and generates re-constructed pictures. The re-constructed pictures are used as reference pictures for motion compensation.

Generally in the methods described above for using temporal references, the prediction references are temporally and spatially as close as possible to the picture, or to the area, which is to be coded. However, predictive coding is vulnerable to transmission errors, since an error affects all pictures that appear in a chain of predicted pictures following that containing the error. Therefore, a

typical way to make a video transmission system more robust to transmission errors is to reduce the length of prediction chains.

Spatial, SNR, and FGS scalability techniques all provide a way to make the critical prediction paths smaller in terms of the number of bytes. A critical prediction path is that part of the bit-stream that needs to be decoded in order to get an acceptable representation of the video sequence contents. In bit-rate-scalable coding, the critical prediction path refers to the base layer of a GOP. It is convenient only to protect the critical prediction path properly rather than the whole one-layer bit-stream.

B-frames can be used instead of temporally corresponding INTER frames in order to shorten prediction paths. However, if the time between consecutive anchor frames is relatively long, the use of B-frames causes a reduction in compression efficiency. In this situation B-frames are predicted from anchor frames which are further away from each other in time and so the B-frames and reference frames from which they are predicted are less similar. This yields a worse predicted B-frame and consequently more bits are required to code the associated prediction error frame. In addition, as the time distance between the anchor frames increases, consecutive anchor frames are less similar. Again, this yields a worse predicted anchor image and more bits are required to code the associated prediction error image.

Temporal prediction normally occurs according to Figure 10.

If the prediction reference of an INTER frame can be selected (as for example in the Reference Picture Selection mode of H.263), prediction paths can be shortened by predicting a current frame from a frame other than the one immediately proceeding it in natural numerical order. This is illustrated in Figure 11.

Video Redundancy Coding (VRC) has been proposed to provide graceful degradation in video quality as a result of packet losses in packet-switched

networks. The principle of VRC is to divide a sequence of pictures into two or more threads in such a way that all pictures are assigned to one of the threads in a round-robin fashion. Each thread is coded independently. At regular intervals, all threads converge into a so-called Sync frame. From this Sync frame, a new thread series is started. The frame rate within one thread is consequently lower than the overall frame rate, half in the case of two threads, a third in the case of three threads and so on. This leads to a substantial coding penalty because of the generally larger differences between consecutive pictures in the same sequence and the longer motion vectors typically required to represent motion-related changes between pictures within a thread. Figure 12 shows VRC operating with two threads and three frames per thread.

If one of the threads is damaged, for example because of a packet loss, it is likely that the remaining threads remain intact and can be used to predict the next Sync frame. It is possible to continue the decoding of the damaged thread, which leads to slight picture degradation, or to stop its decoding which leads to a reduction in the frame rate. If the threads are reasonably short however, both forms of degradation only persist for a very short time, that is until the next Sync frame is reached. The operation of VRC when one of the two threads is damaged is shown in Figure 13.

Sync frames are always predicted from undamaged threads. This means that the number of transmitted INTRA-pictures can be kept small, because there is generally no need for complete re-synchronisation. Correct Sync frame construction is only prevented if all threads between two Sync frames are damaged. In this situation, annoying artefacts persist until the next INTRA-picture is decoded correctly, as would have been the case without employing VRC.

Currently, VRC can be used with ITU-T H.263 video coding standard (version 2) if the optional Reference Picture Selection mode (Annex N) is enabled. However, there are no major obstacles of incorporating VRC into other video compression methods.



Backward prediction of P-frames has also been proposed as a method of shortening prediction chains, as shown in Figure 14. This shows a few consecutive frames of a video sequence. There is an INTRA frame (I1) which is inserted into a coded video sequence as a result of an INTRA frame request or as a result of a periodic INTRA frame refresh operation. An INTRA frame request may be made at a point at which there is a scene cut. After a few INTER coded frames (P2, P3, P4 and P5), another INTRA frame request (or periodic INTRA frame refresh operation) is made. Rather than inserting an INTRA frame immediately after the INTRA frame request (or periodic INTRA frame refresh operation), it is instead inserted after a few temporally predicted frames. The frames between the INTRA frame request and the INTRA frame I1 are predicted backwardly in sequence in INTER format one after the other with I1 as the origin of the prediction chain. The backwardly-predicted INTER frames cannot be decoded before I1 is decoded. Consequently, an initial buffering delay greater than the time between the scene cut and the following INTRA frame is required in order to prevent a pause in playback.

The benefit of this approach can be seen by considering how many frames must be successfully transmitted in order to enable decoding of frame P5. If conventional frame ordering, such as that shown in Figure 15 is used, successful decoding of P5 requires that I1, P2, P3, P4 and P5 are transmitted and decoded correctly. In the method shown in Figure 14, successful decoding of P5 requires that I1, P4 and P5 are transmitted and decoded correctly. In other words, this method provides a greater certainty that P5 will be correctly decoded compared with a method that employs conventional frame ordering and prediction.

Although reference picture selection can be used to deal with temporal error propagation in a video sequence, it decreases compression efficiency.

Conventional SNR and spatial scalability coding as well as FGS coding decrease compression efficiency. Moreover, they require the transmitter to decide how to layer video data during encoding.

Figure 16 shows a video communications system 10 which operates according to the ITU-T H.26L standard based upon test model (TML) TML-3 as modified by current recommendations for TML-4. The system 10 has a transmitter side 12 and a receiver side 14. It should be understood that since the system is equipped for

5 bi-directional transmission and reception, the transmitter and receiver sides 12 and 14 can perform both transmission and reception functions and are interchangeable. The system 10 comprises a video coding layer (VCL) and a network adaptation layer (NAL) with network awareness (that is the NAL is able to adapt the arrangement of data to suit the network). The VCL includes both waveform

10 coding and entropy coding, as well as decoding functionality. The NAL packetises the coded video data into service data units (packets) which are handed to a transport coder before transmission over a channel. The NAL also de-packetises coded video data from service data units received from a transport decoder after transmission over a channel. The NAL is capable of partitioning a video bit-stream

15 into coded block patterns and prediction error coefficients separately from other, more important data such as picture type and motion compensation information.

The main task of the VCL is to code video data in an efficient manner. However, as has been discussed in the foregoing, errors adversely affect efficiently coded

20 data and so some awareness of possible errors is included. The VCL is able to interrupt the predictive coding chain and to take measures to compensate for the occurrence and propagation of errors. There are several ways in which this can be done:

- 25 interrupting the temporal prediction chain by introducing INTRA-frames and INTRA-MBs;
- interrupting error propagation by introducing a slice concept, that is, by restricting motion vector prediction to slice bounds;
- introducing a variable length code which can be decoded independently, for example without adaptive arithmetic coding over frames; and
- 30 fast rate allocation to adapt to varying bit rate channels.

Additionally, the VCL identifies priority classes to support quality of service (QoS) mechanisms in networks.

Encoding schemes include information which describes encoded video frames or pictures. The information is defined as syntax elements. A syntax element is a codeword or a group of codewords having a similar functionality in a coding scheme. The syntax elements are divided into priority classes. The priority class of a syntax element is defined according to its coding and decoding dependencies relative to other classes. Decoding dependencies result from the temporal prediction, spatial prediction and the variable length code. The general rules for defining the priority classes are:

1. If a syntax element A can be decoded correctly without knowledge of a syntax element B and a syntax element B cannot be decoded correctly without knowing the syntax element A, then syntax element A has higher priority than syntax element B.
2. If syntax elements A and B can be decoded independently, the degree of influence on image quality of each syntax element determines its priority class.

Erroneous or missing syntax elements only effect the decoding of syntax elements which are in the current branch away from the root of the dependency tree. Therefore, the impact of syntax elements closer to the root of the tree on decoded image quality is greater than those in lower priority classes.

Priority classes are defined on a frame-by-frame basis. If a slice-based image coding mode is used, some adjustment in the assignment of syntax elements to priority classes is performed.

The dependencies between different syntax elements in the current H.26L test model are shown in Figure 17. The current test model has the following 10 priority classes from Class 1, highest priority, to Class 10, lowest priority:

Class 1: PSYNC, PTYPE: Contains the PSYNC, PTYPE syntax element

Class 2: MB\_TYPE, REF\_FRAME: Contains all MB types and reference frame syntax elements in a frame. For INTRA pictures/frames, this class contains no elements.

Class 3: IPM: Contains INTRA-prediction-Mode syntax element;

Class 4: MVD, MACC: Contains Motion Vectors and Motion accuracy syntax elements (TML-2). For INTRA pictures/frames, this class contains no elements.

Class 5: CBP-Intra: Contains all CBP syntax elements assigned to INTRA-MBs in one frame.

5 Class 6: LUM\_DC-Intra, CHR\_DC-Intra: Contains all DC luminance coefficients and all DC chrominance coefficients for all blocks in INTRA-MBs.

Class 7: LUM\_AC-Intra, CHR\_AC-Intra: Contains all AC luminance coefficients and all AC chrominance coefficients for all blocks in INTRA-MBs.

10 Class 8: CBP-Inter, Contains all CBP syntax elements assigned to INTER-MBs in a frame.

Class 9: LUM\_DC-Inter, CHR\_DC-Inter: Contains the first luminance coefficient of each block and the DC chrominance coefficients of all blocks in INTER-MBs.

15 Class 10: LUM\_AC-Inter, CHR\_AC-Inter: Contains the remaining luminance coefficients and chrominance coefficients of all blocks in INTER-MBs.

The main task of the NAL is to transmit the data contained within the identified priority classes in an optimal way adapted to the underlying network. Therefore, a unique data encapsulation method is defined for each underlying network or type of network. The NAL carries out the following tasks:

1. It maps the data contained in the identified syntax element classes into video packets.
2. It transfers the resulting data packets in a manner adapted to the underlying network.

25 It may also provide error protection mechanisms.

Prioritisation of syntax elements used to code compressed video pictures into different priority classes simplifies adaptation to the underlying network. Networks supporting priority mechanisms obtain particular benefit from prioritisation of syntax elements. Some examples are:

- use of priority methods in IP (such as RVSP);
- use of QoS mechanisms in UMTS;
- use of H.223 Annex C or D; and

use of unequal error protection provided by underlying networks.

Different data/telecommunications networks usually have substantially different characteristics. For example, different packet based networks use protocols that  
 5 employ minimum and maximum packet lengths. Some protocols ensure delivery of data packets in the correct order. Therefore, merging of the data for more than one class into a single data packet or splitting of the data representing a given priority class amongst several data packets is applied as required.

10 In general, the VCL checks, by using the network and the transmission protocols, that a certain class and all classes with higher priority inside a particular frame are identified and have been correctly received, that is without bit errors and being of the correct length.

15 The coded video bit-stream is encapsulated in various ways depending on the underlying network and the application in use. In the following, various encapsulation schemes are presented.

#### H.324 (Circuit-Switched Videophone)

20 The transport coder of H.324, namely H.223, has a maximum service data unit size of 254 bytes. Typically this is not sufficient to carry a whole picture, and therefore the VCL is likely to divide the picture into multiple partitions so that each partition fits into one service data unit. As of currently, codewords have been  
 25 grouped to partitions based on their type (that is the same types of codewords of the picture into the same partition). The codeword (and byte) order of partitions is arranged with decreasing order of importance. If a bit error affects an H.223 service data unit carrying video data, the decoder may lose decoding synchronisation (due to variable length coding of parameters), and the rest of the data in the service data unit cannot be decoded. However, since the most  
 30 important data appeared in the beginning of the service data unit, the decoder is likely to be able to generate a degraded representation of the picture contents.

#### IP Videophone

The maximum size of an IP packet is about 1500 bytes (for historical reasons). It is beneficial to use as large IP packets as possible for two reasons:

1. IP network elements, such as routers, may become congested due to excessive IP traffic, causing internal buffer overflows. The buffers are typically packet-orientated, that is, they can contain a certain number of packets. Thus, in order to avoid network congestion, it is desirable to use rarely generated large packets rather than frequently generated small packets.
2. Each IP packet contains header information. A typical protocol combination used for real-time video communication, namely RTP/UDP/IP, includes a 40-byte header section per packet. A circuit-switched low-bandwidth dial-up link is often used when connecting an IP network. The packetisation overhead becomes significant in low-bit rate links if there are many small packets.

An INTER-coded video picture may comprise sufficiently few bits to fit into a single IP packet (depending on image size and complexity). There are numerous ways to provide unequal error protection in IP networks. These mechanisms include packet duplication, forward error correction (FEC) packets, Differentiated Services (giving priority of a packet in a network), and Integrated Services (RSVP protocol). Typically, these mechanisms require that data with similar importance is encapsulated in one packet.

### IP Video Streaming

As video streaming is a non-conversational application, there are no strict end-to-end delay requirements. Consequently, the packetisation scheme may utilise information from multiple pictures. For example, the data can be classified in a manner similar to the IP videophone case, but with high-importance data from multiple pictures packetised into the same packet.

Alternatively, each picture or slice can be packetised into its own packet. Data partitioning is applied so that the most important data appears at the beginning of packets. Forward Error Correction (FEC) packets are calculated from a set of already transmitted packets. The FEC algorithm is selected so that it protects only a certain number of bytes appearing in the beginning of packets. At the receiving

end, if a normal data packet is lost, the beginning of the lost data packet can be corrected using the FEC packet (as proposed in A. H. Li, J. D. Villasenor, "A generic Uneven Level Protection (ULP) proposal for Annex I of H.323", ITU-T, SG16, Question 15, document Q15-J-61, 16-May-2000).

5

According to a first aspect of the invention there is provided a method for encoding a video signal comprising the steps of:

encoding a first complete frame by forming a bit-stream containing information for its subsequent full re-construction the information being prioritised into high and low priority information;

10

defining at least one virtual frame on the basis of a version of the first complete frame constructed using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

15

encoding a second complete frame by forming a bit-stream containing information for its subsequent full re-construction such that the second complete frame can be fully re-constructed on the basis of the virtual frame rather than on the basis of the first complete frame.

20

In one embodiment, the invention is a scalable coding method. In this case, the virtual frames may be interpreted as a base layer of a scalable bit-stream.

Preferably the method comprises the steps of:

priorising information of the second complete frame into high and low priority information; and

25

encoding a third complete frame by defining a virtual frame on the basis of a version of the third complete frame constructed using high priority information of the third complete frame in the absence of at least some of the low priority information of the third complete frame the third virtual frame being predicted from a version of the second complete frame constructed using the high priority information of the second complete frame in the absence of at least some of the low priority information of the second complete frame. Therefore, a virtual frame

30

can be defined predicted from another virtual frame and refined using its own high priority information. Accordingly, chains of prediction chains may be provided.

5 Preferably the information for the subsequent full re-construction of the complete frame is prioritised into high and low priority information according to its significance in producing a fully re-constructed version of the complete frame.

10 The complete frames may be base layers of a scalable frame. They are complete in the sense that an image capable of display can be formed. This is not necessarily true for the virtual frames.

15 Preferably a plurality of virtual frames are predicted in a prediction chain. The original frame in the chain may be a normal INTRA frame or it may be a virtual INTRA frame.

20 A virtual frame is one which is formed using high priority information and deliberately not using low priority information. Preferably the virtual frame is not displayed. Alternatively, if it is displayed, it is used as an alternative to a complete frame. This may be the case if the complete frame is not available due to a transmission error.

The invention enables improving of the coding efficiency when shortening the temporal prediction path.

25 Preferably the information comprise codewords. Virtual frames may be constructed not exclusively from or defined by high priority information but may also be constructed from or defined by some low priority information.

30 A virtual frame may be predicted from a prior virtual or complete frame. Alternatively or additionally, a virtual frame may be predicted from a subsequent virtual or complete frame using backward-prediction of virtual frames. Backward prediction of INTER frames has been described in the foregoing in relation to



Figure 14. It will be understood that this principle can readily be applied to virtual frames.

Preferably a virtual frame may be decoded using both its high and low priority information (and there may not even have been a division of its information into high and low priority information) and be predicted on the basis of another virtual frame. In this way, a virtual frame may be defined or constructed, even though its information is not prioritised into high and low priority.

Preferably decoding of a bit-stream for a virtual frame uses a different algorithm from decoding of a bit-stream for a complete frame. There may be multiple algorithms for decoding virtual frames. Selection of a particular algorithm may be signalled in the bit-stream.

Preferably in the absence of low priority information, it may be replaced by default values in order to be able to carry out decoding of a virtual frame. The selection of the default values may vary and the correct selection may be signalled in the bit-stream.

According to a second aspect of the invention there is provided a method for decoding a video signal comprising the steps of:

decoding a first complete frame from a bit-stream containing information for its subsequent full re-construction the information being prioritised into high and low priority information;

defining at least one virtual frame on the basis of a version of the first complete frame constructed using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

predicting a second complete frame on the basis of the virtual frame rather than on the basis of the first complete frame.

Preferably the method comprises the step of decoding a third complete frame by defining a virtual frame on the basis of a version of the third complete frame

constructed using high priority information of the third complete frame in the absence of at least some of low priority information of the third complete frame the third virtual frame being predicted from a virtual version of the second complete frame constructed using the high priority information of the second complete frame in the absence of at least some of the low priority information of the second complete frame. Therefore, a virtual frame can be defined predicted from another virtual frame and refined using its own high priority information. Accordingly, chains of prediction chains may be provided. A complete frame may be decoded from a virtual frame. A complete frame may be decoded from a prediction chain of virtual frames.

According to a third aspect of the invention there is provided a video encoder for encoding a video signal comprising:

a complete frame encoder for forming a bit-stream of a first complete frame containing information for subsequent full re-construction of the first complete frame the information being prioritised into high and low priority information;

a virtual frame encoder defining at least one virtual frame as a version of the first complete frame constructed using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

a frame predictor for predicting a second complete frame on the basis of the virtual frame rather than on the basis of the first complete frame.

Preferably the encoder sends a signal to the decoder to indicate which part of the bit-stream for a frame is sufficient to produce an acceptable picture to replace a full-quality picture in case of a transmission error or loss. The signalling may be included in the bit-stream or it may be transmitted separately from the bit-stream. Rather than applying to a frame, the signalling may apply to a part of a picture, for example a slice, a block, a macroblock or a group of blocks. The signalling may indicate which one of multiple pictures may be sufficient to produce an acceptable picture to replace a full-quality picture.

Preferably the encoder can send a signal to the decoder to indicate how to construct a virtual spare reference picture that is used if the actual reference picture is lost or too corrupted.

- 5 According to a fourth aspect of the invention there is provided a decoder for decoding a video signal comprising:
  - a complete frame decoder for decoding a first complete frame from a bit-stream containing information for its subsequent full re-construction the information being prioritised into high and low priority information;
  - 10 a virtual frame decoder for forming at least one virtual frame from the bit stream of the first complete frame using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and
  - a frame predictor for predicting a second complete frame on the basis of the virtual
  - 15 frame rather than the first complete frame.

The invention applies to situations in which the low priority information is not used in the construction of virtual frames because the low priority information has been lost, for example during transmission. Alternatively, the decoder may choose not to

20 use low priority information which it actually possesses. This may be the case if the invention is using Reference Picture Selection and the decoder decides to predict from a reference frame which is not the next naturally occurring frame in a prediction chain. In this way, prediction chains may be broken up. This can deal with the problem of drifting.

25 In the case of Reference Picture Selection, the encoder and the decoder may be provided with multi-frame buffers for storing complete frames and a multi-frame buffer for storing virtual frames.

30 Preferably, a reference frame used to predict another frame may be selected, for example by the encoder, the decoder or both. The selection of the reference frame can be made separately for each frame, picture segment, macroblock, block or whatever sub-picture element. A reference frame can be any complete or virtual

frame that is accessible or that can be generated both in the encoder and in the decoder.

In this way, each complete frame is not restricted to a single virtual frame but may be associated with a number of different virtual frames, each having a different way to classify the bit-stream for the complete frame. These different ways to classify the bit-stream may be different reference (virtual or complete)\_picture(s) for motion compensation and/or a different way of decoding the high priority part of the bit-stream.

Preferably feedback is provided from the decoder to the encoder. This feedback may be in the form of an indication that concerns indicated codewords of one or more specified pictures. The indication may indicate that codewords have been received, have not been received or have been received in a damage state. This may cause the encoder to re-send the codewords. The indication may specify codewords within a certain area within one picture or may specify codewords within a certain area in multiple pictures

According to a fifth aspect of the invention there is provided a video communications system for encoding and decoding a video signal, the system comprising an encoder and a decoder, the encoder comprising:

a complete frame encoder for forming a bit-stream of a first complete frame containing information for subsequent full re-construction of the first complete frame the information being prioritised into high and low priority information;

a virtual frame encoder defining at least one virtual frame as a version of the first complete frame constructed using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

a frame predictor for predicting a second complete frame on the basis of the virtual frame rather than on the basis of the first complete frame;

and the decoder comprising:

a complete frame decoder for decoding a first complete frame from a bit-stream containing information for its subsequent full re-construction the information being prioritised into high and low priority information;

5 a virtual frame decoder for forming at least one virtual frame from the bit stream of the first complete frame using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

a frame predictor for predicting a second complete frame on the basis of the virtual frame rather than the first complete frame.

10

According to a sixth aspect of the invention there is provided a video communications terminal comprising a video encoder, the video encoder comprising:

15 a complete frame encoder for forming a bit-stream of a first complete frame containing information for subsequent full re-construction of the first complete frame the information being prioritised into high and low priority information;

a virtual frame encoder defining at least one virtual frame as a version of the first complete frame constructed using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

20

a frame predictor for predicting a second complete frame on the basis of the virtual frame rather than on the basis of the first complete frame.

25

According to a seventh aspect of the invention there is provided a video communications terminal comprising a decoder, the decoder comprising:

a complete frame decoder for decoding a first complete frame from a bit-stream containing information for its subsequent full re-construction the information being prioritised into high and low priority information;

30

a virtual frame decoder for forming at least one virtual frame from the bit stream of the first complete frame using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

a frame predictor for predicting a second complete frame on the basis of the virtual frame rather than the first complete frame.

According to an eighth aspect of the invention there is provided a computer

5 program for operating a computer as a video encoder comprising:

computer executable code for encoding a first complete frame by forming a bit-stream containing information for its subsequent full re-construction the information being prioritised into high and low priority information;

10 computer executable code for defining at least one virtual frame on the basis of a version of the first complete frame constructed using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

15 computer executable code for encoding a second complete frame by forming a bit-stream containing information for its subsequent full re-construction the information being prioritised into high and low priority information enabling the second complete frame to be fully re-constructed on the basis of the virtual frame rather than on the basis of the first complete frame.

According to an ninth aspect of the invention there is provided a computer

20 program for operating a computer as a video decoder comprising:

computer executable code for decoding a first complete frame from a bit-stream containing information for its subsequent full re-construction the information being prioritised into high and low priority information;

25 computer executable code for defining at least one virtual frame on the basis of a version of the first complete frame constructed using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

30 computer executable code for predicting a second complete frame on the basis of the virtual frame rather than on the basis of the first complete frame.

Preferably the computer programs of the eighth and ninth aspects are storage on a data storage medium. This may be a portable data storage medium or a data

storage medium in a device. The device may be portable, for example a laptop, a personal digital assistant or a mobile telephone.

References to "frames" in the context of the invention is intended also to include parts of frames, for example slices, blocks and MBs, within a frame.

Compared to PFGS, the invention provides better compression efficiency. This is because it has a more flexible scalability hierarchy. It is possible for PFGS and the invention to exist in the same coding scheme. In this case, the invention operates underneath the base layer of PFGS.

The invention enables prediction from virtual frames, that is prediction from the most significant part of the bit-stream. In the prior art, the motion compensation uses only normal frames, that is the entire bit-stream, as references for motion compensation. If there is unequal error protection between the high and low priority information, the invention provides a gain in compression efficiency. Unequal error protection can involve transmission of packets containing high and low priority information, in such a way that the high priority information packets are less likely to be lost. If there is a video bit-stream containing INTRA and INTER frames, to prepare against packet losses, the usage of low priority packets is limited in long prediction chains. As an example of this, reference picture selection can be used to produce a temporally scalable bit-stream as follows: I0 P2 P4 P6 P8 P10(I0) P12 P14 P16 P18 P20(P10) ... , where the value after the frame type corresponds to a time-stamp (temporal reference) and the frame in parentheses is the reference frame for motion compensation. If no reference frame is listed, the previous frame is used as a reference. Frames I0, P10, P20, ... can be encapsulated in high priority packets and the rest of the frames can be encapsulated in low priority packets. The invention provides an advantageous way to utilise information carried in the high priority packets. This may involve encapsulating the high priority parts of low priority frames in high priority packets, for example by transmitting the motion information of P2, P4, P6, and P8 in a high priority packet. The invention provides a method for the encoder to use only this high priority information part of the low priority frames as a prediction reference.

The compression gain is achieved since more data is carried in high priority packets and since the new coding scheme allows utilisation of all data carried in high priority packets (even though the data may not contain all data necessary for a fully re-constructable frame). The invention provides (virtual) reference frames that resemble the frames which are to be predicted more than the prior art. Consequently, the invention needs fewer bits to encode the prediction error information.

The invention will now be described, by way of example only, with reference to the accompanying drawings in which:

Figure 1 shows a video transmission system;

Figure 2 shows B-pictures being predicted from anchor picture pairs;

Figure 3 shows an IP multicasting system;

Figure 4 shows SNR scalable pictures;

Figure 5 shows spatial scalable pictures;

Figure 6 shows prediction relationships in fine granularity scalable coding;

Figure 7 shows conventional prediction relationships used in scalable coding;

Figure 8 shows prediction relationships in progressive fine granularity scalable coding;

Figure 9 illustrates channel adaptation in progressive fine granularity scalability;

Figure 10 shows conventional temporal prediction;

Figure 11 illustrates the shortening of prediction paths using Reference Picture Selection;

Figure 12 illustrates Video Redundancy Coding;

Figure 13 shows Video Redundancy Coding dealing with damaged threads;

Figure 14 illustrates re-positioning an INTRA frame and backward prediction of INTER frames;

Figure 15 shows conventional frame prediction relationships following an INTRA-frame;

Figure 16 shows a video transmission system;

Figure 17 shows dependencies of syntax elements in the H.26L TML-4 test model;

Figure 18 shows an encoding procedure;

Figure 19 shows a decoding procedure;



Figure 20 shows a modification of the decoding procedure of Figure 19;  
Figure 21 illustrates a video coding method according to the invention;  
Figure 22 illustrates another video coding method according to the invention;  
Figure 23 shows a video transmission system according to the invention; and  
5 Figure 24 shows a video transmission system utilising ZPE-pictures.

Figures 1 to 17 have been described in the foregoing.

10 The invention may be implemented in a video transmission system according to Figure 16.

In order to describe the invention, it will firstly be described as an algorithm to illustrate its principles and then will be described in greater detail with reference to Figures 18, 19 and 20. The algorithm has a first phase which occurs during  
15 encoding and a second phase which occurs during decoding.

In the first phase (packetisation), a modified bit-stream containing only the signalled parameters is generated. This may be obtained directly from the NAL or it can be created using the algorithm. The algorithm first parses the bit-stream  
20 corresponding to the last picture that should be used for reference generation. Then it generates another representation of the bit-stream that includes only the signalled parameter classes and uses default values for non-signalled parameters (zero, not existing, etc.). At the same time, it determines which frames are used as prediction references for the parsed frame. If a reference frame is not a virtual one  
25 and if it is in the signalled range of frames, the algorithm recursively parses the bit-stream for the reference frame.

In the second phase, the algorithm decodes the newly generated bit-stream starting from the bit-stream section that corresponds to the signalled first picture used for prediction. If the bit-stream refers to a reference frame which does not  
30 exist in a virtual multi-frame reference buffer, it is obtained from a normal multi-frame reference buffer. Otherwise, the algorithm uses a re-constructed frame from the virtual multi-frame reference buffer. Each re-constructed frame is placed in the

virtual multi-frame reference buffer. Finally, the output of the algorithm resides in the virtual multi-frame reference buffer.

The algorithm requires there to be a similar but separate multi-frame reference  
5 buffer which is used for normal decoding. Each entry in the buffer should be capable of storing the re-constructed frame and the bit-stream for the frame. A multi-picture buffer as described in H.263 Draft Annex U (having appropriate changes to H.26L) can be used.

10 The invention will now be described in greater detail as a set of procedural steps with reference to Figure 18 which illustrates an encoding procedure carried out by an encoder and Figure 19 which illustrates a decoding procedure carried out by a decoder corresponding to the encoder.

15 Referring now to the procedure of Figure 18. In an initialisation phase, the encoder initialises a frame counter (step 110), initialises a normal reference frame buffer (step 112) and initialises a virtual reference frame buffer (step 114). The encoder then receives raw uncoded video data from a source (step 116), such as a video  
20 camera. The video data may originate from a live feed. The encoder receives an indication of the coding mode to be used in the coding of a current frame (step 118), that is, whether it is to be an INTRA frame or an INTER frame. The indication can come from a pre-set coding scheme (block 120). The indication can optionally come from a scene cut detector (block 122), if one is provided, or as feedback from a decoder (block 124). The encoder then makes a decision whether to code  
25 the current frame as an INTRA frame (step 126).

If the decision is "YES" (decision 128), the current frame is encoded to form a compressed frame in INTRA frame format (step 130).

30 If the decision is "NO" (decision 132), the encoder receives an indication of a frame to be used as a reference in encoding the current frame in INTER frame format (step 134). This can be determined as a result of a predetermined coding scheme (block 136). In another embodiment of the invention, this may be

controlled by feedback from the decoder (block 138). This will be described later. The identified reference frame may be a normal frame or a virtual frame and so the encoder determines whether a virtual reference is to be used (step 140).

- 5 If a virtual reference frame is to be used, it is retrieved from the virtual reference frame buffer (step 142). If a virtual reference is not to be used, a normal reference frame is retrieved from the normal frame buffer (step 144). This pre-supposes the presence of normal and virtual reference frames in their respective buffers. If the encoder is transmitting the first frame following initialisation, this is usually an
- 10 INTRA frames and so no reference frame is used. The current frame is then encoded into INTER frame format using the raw video data and the selected reference frame (step 146).

- 15 Irrespective of whether the current frame is encoded into INTRA frame format or INTER frame format, the following steps are then applied. The encoded frame data are prioritised (step 148), the particular prioritisation depending on whether INTER frame or INTRA frame coding has been used. The prioritisation divides the data into low priority and high priority data on the basis of how essential it is to the re-construction of the picture being encoded. Once so divided, a bit-stream is
- 20 formed for transmission. In forming the bit-stream, a suitable packetisation method is used such as those referred to above. The bit-stream is then transmitted to the decoder (step 152). If the current frame is the last frame, a decision is made (step 154) to terminate the procedure (block 156) at this point.

- 25 Assuming that the current frame is not the last frame, the bit-stream representing the current frame is decoded on the basis of the relevant reference frame using both the low priority and high priority data in order to form a complete re-construction of the frame (step 157). The complete re-construction is then stored in the normal reference frame buffer. The bit-stream representing the current
- 30 frame is then decoded on the basis of the relevant reference frame using only the high priority data in order to form a re-construction of a virtual frame (step 160). The re-construction of the virtual frame is then stored in the virtual reference frame

buffer (step 162). The set of procedural steps starts again from step 116 and the next frame is then encoded and formed into a bit-stream.

5 The order of the steps presented above may be different in an alternative embodiment of the invention. For example, the initialisation steps can occur in any convenient order, as can the steps of decoding the complete re-construction of the normal reference frame and the re-construction of the virtual reference frame.

10 Although the foregoing describes a frame being predicted from a single reference, in another embodiment of the invention, more than one reference frame can be used to predict one picture. The selection of the reference frame can be made for each picture segment, macroblock, block or whatever sub-picture separately of the frame being encoded/decoded. A reference frame can be any complete or virtual frame that is accessible or can be generated both in the encoder and in the  
15 decoder. In some cases, such as B frames, two or more reference frames are associated for the same picture area, and some kind of interpolation scheme is used to predict the area to be coded. Each complete frame may be associated with a number of different virtual frames, representing:

different ways to classify the bit-stream for the complete frame;  
20 different reference (virtual or complete) pictures for motion compensation; and/or different ways of decoding the high priority part of the bit-stream.

Therefore, in such an embodiment, references to normal and virtual reference frame buffers would in fact be references to normal and virtual reference frame buffers.

25 Reference will now be made to the procedure of Figure 19. In an initialisation phase the decoder initialises a virtual reference frame buffer (step 210), a normal reference frame buffer (step 211) and a frame counter (step 212). The decoder then receives a bit-stream relating to a compressed current frame (step 214). The  
30 decoder then determines whether the current frame is encoded in INTRA frame format or INTER frame format (step 216). This can be determined from information received, for example, in the picture header.

If the current frame is in INTRA frame format, it is decoded using the complete bit-stream to form a complete re-construction of the INTRA frame (step 218).

- 5 If the current frame is the last frame then a decision is made (step 220) to terminate the procedure (step 222). Assuming the current frame is not the last frame, the bit-stream representing the current frame is then decoded using high priority data in order to form a virtual reference frame (step 224).

- 10 If the current frame is in INTER frame format, the reference frame used in its prediction (at the encoder) is identified (step 226). The reference frame may be identified by data present in the transmitted bit stream.

The identified reference may be a normal frame or a virtual frame and so the decoder determines whether a virtual reference is to be used (step 228).

15

If a virtual reference is to be used, it is retrieved from the virtual reference frame buffer (step 230). If it is not, a normal reference frame is retrieved from the normal reference frame buffer (step 232). It should be noted that decoding of the high priority information to construct a virtual frame does not follow the same decoding procedure as decoding the complete representation of the frame. For example, in the absence of low priority information, it may be replaced by some default values in order to be able to carry out decoding of the virtual frame.

20

This pre-supposes the presence of normal and virtual reference frames in their respective buffers. If the decoder is receiving the first frame following initialisation, this is usually an INTRA frames and so no reference frame is used.

25

The current frame is then decoded in INTER frame format using the complete received bit-stream and the identified reference frame as a prediction reference (step 234).

30

If the current frame is the last frame then a decision is made (step 236) to terminate the procedure (step 222). Assuming that the current frame is not the last

frame, the bit-stream representing the current frame is then decoded using high priority data in order to form a virtual reference frame (step 238). This virtual reference frame is then stored in the virtual reference frame buffer (step 240).

- 5 As mentioned in the foregoing, in one embodiment of the invention, selection of a normal or a virtual reference in the encoder is carried out on the basis of feedback from the decoder.

10 Figure 20 shows additional steps to modify the procedure of Figure 19 to provide this feedback. The additional steps of Figure 20 are inserted between steps 214 and 216 of Figure 19. Since Figure 19 has been fully described in the foregoing only the additional steps will be described. This is just one implementation among several possibilities. For example, instead of requesting an INTRA picture, the  
15 decoder could indicate that all of the data for the frame was lost, and the encoder should react to this indication so that it does not refer to the lost frame in motion compensation. In other words, an INTRA picture update is not necessarily needed.

Once a bit-stream for a compressed current frame has been received (step 214), the decoder checks (step 310) whether the bit-stream has been correctly received.  
20 This involves general error checking followed by more specific checks depending on the severity of the error. If the bit-stream has been correctly received then the decoder determines whether the current frame is encoded in INTRA frame format or in INTER frame format (step 216).

25 If the bit-stream has not been correctly received the decoder determines whether it is able to decode the picture header (step 312). If it cannot, it issues an INTRA frame up-date request to the sending terminal containing the encoder (step 314) and the procedure returns to step 214.

30 If the decoder can decode the frame header, it determines whether it is able to decode the high priority data (step 316). If it cannot, step 314 occurs.

If the decoder can decode the high priority data, it determines whether it is able to decode the low priority data (step 318). If it cannot, it instructs the sending terminal containing the encoder to encode the next frame predicted only on the high priority data of the current frame and not on the low priority data (step 320). Therefore, in

5 the invention, there is a new type of an indication which concerns indicated codewords of one or more specified pictures. The indication may indicate codewords which have, and have not, been received. This provides the feedback referred to above in relation to block 138. In this way the next frame is encoded on a virtual reference frame based on the current frame. If the decoder can decode

10 the low priority data the procedure returns to step 214. This applies if there is such a low delay that the encoder can get the feedback information before encoding the next frame. If this is not the case, it is preferred to send an indication that the low priority part of the particular frame was lost. The encoder then reacts to this indication so that it does not use the low priority information in the next frame it is

15 going to encode. In other words, the encoder generates a virtual frame whose prediction chain does not include the lost low priority part.

The procedures of Figure 18 and Figures 19 and 20 can be implemented in the form of a suitable algorithm or suitable algorithms.

20

Decoding of a bit-stream for virtual frames may use a different algorithm from decoding of a bit-stream for complete or normal frames. In one embodiment of the invention, there is a plurality of such algorithms, and the selection of the correct algorithm may be signalled in the bit-stream. In the absence of low priority

25 information, it may be replaced by some default values in order to be able to carry out decoding of a virtual frame. The selection of the default values may vary, and the correct selection may be signalled in the bit-stream for example by using the indication referred to in the preceding paragraph.

30

The procedures of Figures 18, 19 and 20 use a frame-by-frame approach to decoding. In other embodiments of the invention it can be block-by-block or slice-by-slice. In general, the invention can apply to any picture segment, not just slices and blocks.

Although the coding of INTER frames has been described, coding of B-frames has not been described for the sake of simplicity. However, coding of B-frames could be included in an embodiment of the invention. Furthermore, secondary representations of pictures (that is sync frames), can be included in an embodiment of the invention. If a virtual frame is used to predict a sync frame, the decoder does not need to generate it if the primary representation is correctly received. In this case it is not necessary to form a virtual reference frame for other copies of the sync frame. In conventional scalability approached it is always necessary to decode the base layer.

In one embodiment, a video frame is encapsulated in at least two packets, one with high importance and the other one with low importance. If H.26L is used, the low importance packet can contain coded block patterns and prediction error coefficients, for example.

In Figures 18, 19 and 20, reference is made (in blocks 160, 224, 238) to decoding a frame by using high priority information in order to form a virtual frame. This can actually be carried out in two sub-steps:

- 1) generating a temporary bit-stream representation of a frame consisting of the high priority information and default values for the low priority information.
- 2) decoding the temporary representation normally (as in the case where all information is available).

This is just one embodiment of the invention since the selection of default values can be tuned, and since the decoding algorithm of the virtual frame may not be the same as for the normal frame.

Figures 18 and 19 relate to a particular embodiment of the invention. There is no specific limit to the number of virtual frames which can be generated from each normal frame. In a preferred embodiment of the procedure, there are multiple different virtual frame chains since the virtual frames belonging to the prediction chain are generated when needed.



The bit-stream syntax of the invention is very close to the one-layer coding in which scalability enhancement layers are not required. Moreover, since virtual frames are not displayed, encoders can decide how to generate the virtual prediction reference when they start to encode a frame based on it. In other words, encoders can utilise the bit-stream of previous frames flexibly and frames can be divided into different combinations of codewords after they are transmitted. The layering information of previous frames, that is the information which indicates which codewords belong to high priority information, is transmitted when the virtual prediction frame is generated. In the prior art, encoders have to choose the layering division of a frame while encoding the frame and the information is transmitted within the bit-stream of the corresponding frame.

Figure 21 shows in graphical form decoding of a section of a video sequence including INTRA-coded frame I0 and INTER-coded frames P1, P2, and P3. This Figure is provided to show the effect of the procedure described in relation to Figures 19 and 20. As can be seen in Figure 21, there is a top row, a middle row and a bottom row. The top row corresponds to re-constructed and displayed frames, the middle row corresponds to the bit-stream for each frame and the bottom row corresponds to virtual prediction reference frames which are generated. Arrows indicate the input sources used to produce a certain re-constructed virtual frame. For example, frame I0 is generated from a corresponding bit-stream I0 B-S and frame P1 is re-constructed using frame I0 as a motion compensation reference and the received bit-stream for P1. Similarly, virtual frame I0' is generated from a part of the bit-stream corresponding to frame I0 and artificial frame P1' is generated using I0' as a motion compensation reference and a part of the bit-stream for P1. According to the invention, frame P3 is generated using virtual frame P2' as a motion compensation reference and the bit-stream for P3. Frame P2 is not used as a motion compensation reference.

There are circumstances in which separating the data into high and low priority is not necessary. For example, if the whole data relating to a picture can fit into a single frame, then it may be preferred not to separate the data. In this case, the whole data may be used in prediction from a virtual frame. Referring to Figure 21.

In this particular embodiment, frame P1' is constructed by predicting from virtual frame I0' and by decoding all of the bit-stream information for P1. The reconstructed virtual frame P1' is not equivalent to frame P1, because the prediction reference for frame P1 is I0 whereas the prediction reference for frame P1' is I0'.

- 5 Thus, P1' is a virtual frame, even though, in this case, it is predicted from a frame (P1) having information which is not prioritised into high and low priority.

An embodiment of the invention will now be described with reference to Figure 22 as applied to a situation corresponding to that shown in Figure 11. Motion and header data is separated from prediction error data. The motion and header data is encapsulated in a transmission packet called a motion packet and the prediction error data is encapsulated in a transmission packet called a prediction error packet. This is done for several consecutive coded pictures. Motion packets have high priority and they are re-transmitted whenever it is possible and necessary, since error concealment is better if the decoder receives motion information correctly. (Motion packets also improve compression efficiency.) The encoder separates motion and header data from P-frames 1 to 3 and forms a motion packet (M1-3) from that information. Prediction error data for P-frames 1 to 3 is transmitted in a separate prediction error packet (PE1, PE2, PE3). Instead of using I1 as a motion compensation reference, the encoder generates artificial frames P1', P2' and P3' based on I1 and M1-3. In other words, the encoder decodes I1 and the motion part of prediction frames P1, P2, and P3 so that P2' is predicted from P1' and P3' is predicted from P2'. Frame P3' is then used as a motion compensation reference for frame P4. Artificial frames P1', P2' and P3' are referred to as a Zero-Prediction-Error (ZPE) frames since they do not contain any prediction error data.

It is to be noted that a frame and its virtual counterpart are decoded using different parts of the available bit-stream. Normal frames (that is those which are displayed) use all of the available bit-stream and the virtual frames only use part of the bit-stream. The part the virtual frames use is a part of the bit-stream which is most significant in decoding a frame. In addition, it is preferred that the part the virtual frames use is the most robustly protected against errors for transmission, and thus

most likely to be successfully transmitted. In this way, the invention is able to shorten the predictive coding chain and base a predicted frame on an virtual motion compensation reference frame which is generated from the most significant part of a bit-stream rather than on a motion compensation reference which is generated by using the most significant part and a less significant part.

When the procedures of Figure 18, 19 and 20 are applied to H.26L, pictures are encoded containing picture headers. This is the highest priority information referred to in the classification scheme referred to above because without the picture header, the entire picture cannot be decoded. Each picture header contains a picture type (Ptype) field. A particular value is included to indicate whether the picture uses one or more virtual reference pictures. If the value of the Ptype field indicates that one or more virtual reference pictures are to be used, the picture header contains information on how to generate the reference picture. In other embodiments of this invention, this information may be included in slice headers, MB headers and/or block headers depending on what kind of packetisation is used. Furthermore, if multiple reference pictures are used in connection with the encoding of a given frame, one or more of the reference pictures may be virtual. The following signalling schemes are used:

1. An indication of which frames of the past bit-stream is used to generate a reference picture is provided in the transmitted bit-stream. Two values are transmitted: one that corresponds to the temporally last picture and another one that corresponds to the temporally earliest picture that is used for prediction. The procedure of Figures 18 and 19 would have to be suitably modified to use this indication.
2. An indication of which coding parameters are used to generate a reference picture. The bit-stream should carry an indication of the largest priority class that is used for prediction. For example, if the bit-stream carries an indication for class 4, the prediction is formed from parameters belonging to classes 1, 2, 3, or 4. (This is a simplification of a more general scheme where the used classes should be signalled one by one.

Figure 23 shows a video transmission system 400 according to the invention. The system comprises communicating video terminals 402 and 404. In this embodiment, terminal to terminal communication is shown. In another embodiment, there may be a terminal to server or a server to terminal communication configuration. Although it is intended that the system 400 enables bi-directional sending of video data in the form of a bit-stream, it may enable only uni-directional sending of video data. For the sake of simplicity, in the system 400 shown in Figure 24, the video terminal 402 is a sending (encoding) video terminal and the video terminal 404 is a receiving (decoding) video terminal.

The video terminal 402 comprises an encoder 410 and a transceiver 412. The encoder 410 comprises a complete frame encoder 414, a virtual frame encoder 416 and a frame predictor 418. In addition, the encoder comprises a multi-frame buffer 420 for storing normal frames and a multi-frame buffer 422 for storing virtual frames.

The complete frame encoder 414 forms a bit-stream of a normal frame which contains information for its subsequent full re-construction. This carries out the steps 118 to 146 and step 150 of Figure 18. However, the branch containing step 142 is not followed and this step is not used. The information is prioritised into high and low priority information according to step 148 of Figure 18.

The virtual frame encoder 416 defines a virtual frame as a version of the normal frame constructed using the high priority information of the normal frame in the absence of at least some of the low priority information of the first complete frame. This carries out steps 160 and 162 of Figure 18.

The frame predictor 418 encodes another normal frame by forming a bit-stream containing information for its subsequent full re-construction. This carries out the steps 118 to 146 and step 150 of Figure 18. However, the branch containing step 144 is not followed and this step is not used. Therefore, the encoding of this other normal frame is carried out according to steps 140 and 142 such that it can be fully

re-constructed according to step 146 on the basis of a virtual reference frame rather than on the basis of the normal frame.

5 The video terminal 404 comprises a decoder 423 and a transceiver 424. The decoder 423 comprises a complete frame decoder 425, a virtual frame decoder 426 and a frame predictor 428. In addition, the encoder comprises a multi-frame buffer 430 for storing normal frames and a multi-frame buffer 432 for storing virtual frames.

10 The complete frame decoder 425 decodes a normal frame from a bit-stream containing information for its subsequent full re-construction. This carries out step 218 and step 226 to 234 of Figure 19. The information of the bit-stream had previously been prioritised into high and low priority information.

15 The virtual frame decoder 426 forms a virtual frame from the bit stream of the first complete frame using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame according to steps 224, 238 and 240 of Figure 19.

20 The frame predictor 428 predicts another normal frame according to step 234 of Figure 19 on the basis of the virtual frame rather than the normal frame.

25 The video terminal 402 produces an encoded video bit-stream 434 which is transmitted by the transceiver 412 and received by the transceiver 424 across a suitable transmission medium. In one embodiment of the invention, the transmission medium is an air interface in a wireless communications system. The transceiver 424 transmits feedback 436 to the transceiver 412. The nature of this feedback has been described in the foregoing.

30 Operation of a video transmission system 500 utilising ZPE frames will now be described. The system 500 is shown in Figure 24. The system 500 has a transmitting terminal 510 and a plurality of receiving terminals 512 (only one of which is shown) which communicate over a transmission channel or network. The

transmitting terminal 510 comprises an encoder 514, a packetiser 516 and a transmitter 518. It also comprises a TX-ZPE-decoder 520. The receiving terminals 512 each comprise a receiver 522, a de-packetiser 524 and a decoder 526. They also each comprise a RX-ZPE-decoder 528. The encoder 514 codes  
 5 uncompressed video to form compressed video pictures. The packetiser 516 encapsulates compressed video pictures into transmission packets. It may reorganise the information obtained from the encoder. It also outputs video pictures that contain no prediction error data for motion compensation (called the ZPE-bit-stream). The TX-ZPE-decoder 520 is a normal video decoder that is used  
 10 to decode the ZPE-bit-stream. The transmitter 518 delivers packets over the transmission channel or network. The receiver 522 receives packets from the transmission channel or network. The de-packetiser 524 de-packetises the transmission packets and generates compressed video pictures. If some packets are lost during transmission, the de-packetiser 524 tries to conceal the losses in  
 15 the compressed video pictures. In addition, the de-packetiser 524 outputs the ZPE-bit-stream. The decoder 526 re-constructs pictures from the compressed video bit-stream. The RX-ZPE-decoder 528 is a normal video decoder that is used to decode a ZPE-bit-stream.

20 The encoder 514 operates normally except for the case when the packetiser 516 requests a ZPE frame to be used as a prediction reference. Then the encoder 514 changes the default motion compensation reference picture to the ZPE frame that is delivered by the TX-ZPE-decoder 520. Moreover, the encoder 514 signals the usage of the ZPE frame in the compressed bit-stream, for example in the picture  
 25 type of the picture.

The decoder 526 operates normally except for the case when the bit-stream contains a ZPE frame signal. Then the decoder 526 changes the default motion compensation reference picture to the ZPE frame that is delivered by the RX-ZPE-decoder 528.  
 30

Performance of the invention is presented compared against reference picture selection as specified in the current H.26L recommendation. Three commonly

available test sequences are compared, namely Akiyo, Coastguard, and Foreman. The resolution of the sequences is QCIF, having a luminance picture size of 176 x 144 pixels and a chrominance picture size of 88 x 72 pixels. Akiyo and Coastguard are captured with 30 frames per second, whereas the frame rate of Foreman is 25 frames per second. The frames were coded with an encoder following ITU-T recommendation H.263. In order to compare different methods, a constant target frame rate (of 10 frames per second) and a number of constant image quantisation parameters were used. The thread length,  $L$ , was selected so that the size of the motion packet was less than 1400 bytes (that is, that the motion data for a thread was less than 1400 bytes).

The ZPE-RPS case has frames  $I1, M1-L, PE1, PE2, \dots, PEL, P(L+1)$  (predicted from  $ZPE1-L, P(L+2), \dots$ ), whereas the normal RPS case has frames  $I1, P1, P2, \dots, PL, P(L+1)$  (predicted from  $I1, P(L+2)$ ). The only frame coded differently in the two sequences was  $P(L+1)$ , but the image quality of this frame in both sequences is similar due to use of a constant quantisation step. The table below shows the results:

	QP	Number of coded frames in thread, $L$	Original bit rate (bps)	Bit rate increase, ZPE-RPS (bps)	Bit rate increase, ZPE-RPS (%)	Bit rate increase, normal RPS (bps)	Bit rate increase, normal RPS (%)
Akiyo	8	50	17602	14	0.1%	158	0.9%
	10	53	12950	67	0.5%	262	2.0%
	13	55	9410	42	0.4%	222	2.4%
	15	59	7674	-2	0.0%	386	5.0%
	18	62	6083	24	0.4%	146	2.4%
	20	65	5306	7	0.1%	111	2.1%
Coastguard	8	16	107976	266	0.2%	1505	1.4%
	10	15	78458	182	0.2%	989	1.3%

	15	15	43854	154	0.4%	556	1.3%
	18	15	33021	187	0.6%	597	1.8%
	20	15	28370	248	0.9%	682	2.4%
Foreman	8	12	87741	173	0.2%	534	0.6%
	10	12	65309	346	0.5%	622	1.0%
	15	11	39711	95	0.2%	266	0.7%
	18	11	31718	179	0.6%	234	0.7%
	20	11	28562	-12	0.0%	-7	0.0%

It can be seen from the bit-rate increase columns of the results that Zero-Prediction-Error frames improve the compression efficiency when Reference Picture Selection is used.

5

A high-importance part of the bit-stream can be re-constructed and used to conceal loss or corruption of a low-importance part of the bit-stream. One aspect of the invention is to control error concealment in an explicit way. This occurs by the encoder signalling to the decoder which part of the bit-stream for a frame is sufficient to produce an acceptable picture to replace a full-quality picture in case of a transmission error or loss. The signalling can be included in the video bit-stream or it can be transmitted separately from the video bit-stream. The decoder decodes the high-importance part and replaces the low-importance part by default values, in order to get an acceptable picture to display. The principle can be applied to sub-pictures (slices etc.) and also to multiple pictures. In another error concealment approach, the encoder can signal to the decoder how to construct a virtual artificial spare reference picture that is used if the actual reference picture is lost or becomes too corrupted to be readily used.

10

15

20

It should be noted that the virtual prediction reference frames do not necessarily represent contents of any uncompressed picture appearing in the sequence. In known scalability techniques, the prediction reference pictures are a representation of the corresponding original picture. Since the virtual prediction reference frames are not intended to be displayed, unlike the base layer in the



traditional scalability schemes, it is not necessary for the encoder to construct an acceptable quality virtual pictures. Consequently the compression efficiency of the invention is close to a one-layer coding approach.]

- 5 It is possible to extract or decode the same information from the bit-stream in different ways to enable construction of different kinds of virtual frames.

The invention improves the coding efficiency in reference picture selection schemes. The invention can be classified as a new type of SNR scalability that is  
10 more flexible than prior art scalability techniques.

Particular implementations and embodiments of the invention have been described. It is clear to a person skilled in the art that the invention is not restricted to details of the embodiments presented above, but that it can be implemented in  
15 other embodiments using equivalent means without deviating from the characteristics of the invention. The scope of the invention is only restricted by the attached patent claims.

## Claims

1. A method for encoding a video signal comprising the steps of:

encoding a first complete frame by forming a bit-stream containing information for its subsequent full re-construction (150) the information being prioritised (148) into

5 high and low priority information;

defining (160) at least one virtual frame on the basis of a version of the first complete frame constructed using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

10 encoding (146) a second complete frame by forming a bit-stream containing information for its subsequent full re-construction such that the second complete frame can be fully re-constructed on the basis of the virtual frame rather than on the basis of the first complete frame.

15 2. A method according to claim 1 comprising the steps of:

priorising information of the second complete frame into high and low priority information; and

encoding (146) a third complete frame by defining a virtual frame on the basis of a version of the third complete frame constructed using high priority information of the third complete frame in the absence of at least some of the low priority information of the third complete frame the third virtual frame being predicted from a version of the second complete frame constructed using the high priority information of the second complete frame in the absence of at least some of the low priority information of the second complete frame.

25 3. A method according to claim 1 or claim 2 comprising the step of changing a temporal prediction path by predicting on the basis of the virtual frame (142) rather than on the basis of the first complete frame (144).

30 4. A method according to any preceding claim comprising the step of selecting a particular reference frame amongst a plurality of choices to predict another frame.

5. A method according to any preceding claim comprising the step of associating each complete frame with a plurality of different virtual frames, each representing a different way to classify the bit-stream for the complete frame.

5 6. A method according to any preceding claim comprising the step of encoding a virtual frame using both its high and low priority information and predicting it on the basis of another virtual frame.

7. A method according to any preceding claim comprising the step of encoding  
10 virtual frames by using multiple algorithms.

8. A method according to claim 7 comprising the step of signalling in the bit-stream the selection of a particular algorithm.

15 9. A method according to any preceding claim comprising the step of replacing low priority information by default values in order to be able to carry out decoding of a virtual frame.

10. A method for decoding a video signal comprising the steps of:

20 decoding (218) a first complete frame from a bit-stream containing information for its subsequent full re-construction the information being prioritised into high and low priority information;

defining (224) at least one virtual frame on the basis of a version of the first complete frame constructed using the high priority information of the first complete  
25 frame in the absence of at least some of the low priority information of the first complete frame; and

predicting (234) a second complete frame on the basis of the virtual frame (230) rather than on the basis of the first complete frame (232).

30 11. A method according to claim 10 comprising the step of decoding (234) a third complete frame by defining a virtual frame on the basis of a version of the third complete frame constructed using high priority information of the third complete frame in the absence of at least some of low priority information of the third

complete frame the third virtual frame being predicted from a virtual version of the second complete frame constructed using high priority information of the second complete frame in the absence of at least some low priority information of the second complete frame.

5

12. A method according to any preceding claim comprising the step of prioritising (148) the information for the subsequent full re-construction of the complete frame into high and low priority information according to its significance in producing a fully re-constructed version of the complete frame.

10

13. A video encoder (410) for encoding a video signal comprising:

a complete frame encoder (414) for forming a bit-stream of a first complete frame containing information for subsequent full re-construction of the first complete frame the information being prioritised into high and low priority information;

15

a virtual frame encoder (416) defining at least one virtual frame as a version of the first complete frame constructed using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

a frame predictor (418) for encoding a second complete frame by forming a bit-stream containing information for its subsequent full re-construction such that the second complete frame can be fully re-constructed on the basis of the virtual frame rather than on the basis of the first complete frame.

20

14. An encoder (410) according to claim 13 which sends a signal to a corresponding decoder to indicate which part of the bit-stream for a frame is sufficient to produce an acceptable picture to replace a full-quality picture in case of a transmission error or loss of information.

25

15. An encoder (410) according to claim 14 in which the signal indicates which one of multiple pictures is sufficient to produce an acceptable picture to replace a full-quality picture.

30

16. An encoder (410) according to any of claims 13 to 15 which is provided with a multi-frame buffer for storing complete frames (420) and a multi-frame buffer for storing virtual frames (422).

- 5 17. A decoder (423) for decoding a video signal comprising:  
 a complete frame decoder (425) for decoding a first complete frame from a bit-stream containing information for its subsequent full re-construction the information being prioritised into high and low priority information;  
 a virtual frame decoder (426) for forming at least one virtual frame from the bit  
 10 stream of the first complete frame using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and  
 a frame predictor (428) for predicting a second complete frame on the basis of the virtual frame rather than the first complete frame.

15

18. A decoder according to claim 17 which is provided with a multi-frame buffer for storing complete frames (430) and a multi-frame buffer for storing virtual frames (432).

- 20 19. A decoder according to claim 17 or claim 18 in which feedback (436) is provided from the decoder to a corresponding encoder in the form of an indication that concerns indicated codewords of one or more specified pictures.

- 25 20. A video communications terminal (402) comprising a video encoder (410), the video encoder comprising:  
 a complete frame encoder (414) for forming a bit-stream of a first complete frame containing information for subsequent full re-construction of the first complete frame the information being prioritised into high and low priority information;  
 a virtual frame encoder (416) defining at least one virtual frame as a version of the  
 30 first complete frame constructed using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

a frame predictor (418) for encoding a second complete frame by forming a bit-stream containing information for its subsequent full re-construction such that the second complete frame can be fully re-constructed on the basis of the virtual frame rather than on the basis of the first complete frame.

5

21. A video communications terminal (404) comprising a decoder (423), the decoder comprising:

a complete frame decoder (425) for decoding a first complete frame from a bit-stream containing information for its subsequent full re-construction the  
10 information being prioritised into high and low priority information;

a virtual frame decoder (426) for forming at least one virtual frame from the bit stream of the first complete frame using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

15 a frame predictor (428) for predicting a second complete frame on the basis of the virtual frame rather than the first complete frame.

22. A computer program for operating a computer as a video encoder comprising:

20 computer executable code for encoding a first complete frame by forming a bit-stream containing information for its subsequent full re-construction the information being prioritised into high and low priority information;

computer executable code for defining at least one virtual frame on the basis of a version of the first complete frame constructed using the high priority information of the first complete frame in the absence of at least some of the low priority  
25 information of the first complete frame; and

computer executable code for encoding a second complete frame by forming a bit-stream containing information for its subsequent full re-construction such that the second complete frame to be fully re-constructed on the basis of the virtual frame rather than on the basis of the first complete frame.

30

23. A computer program for operating a computer as a video decoder comprising:

computer executable code for decoding a first complete frame from a bit-stream containing information for its subsequent full re-construction the information being prioritised into high and low priority information;

computer executable code for defining at least one virtual frame on the basis of a

- 5 version of the first complete frame constructed using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

computer executable code for predicting a second complete frame on the basis of the virtual frame rather than on the basis of the first complete frame.

## Abstract

A method for encoding a video signal comprises the steps of:

encoding a first complete frame by forming a bit-stream containing information for its subsequent full re-construction (150) the information being prioritised (148) into high and low priority information;

defining (160) at least one virtual frame on the basis of a version of the first complete frame constructed using the high priority information of the first complete frame in the absence of at least some of the low priority information of the first complete frame; and

encoding (146) a second complete frame by forming a bit-stream containing information for its subsequent full re-construction the information being prioritised into high and low priority information enabling the second complete frame to be fully re-constructed on the basis of the virtual frame rather than on the basis of the first complete frame. A corresponding decoding method is also described.

(Figs 18a & 18b)



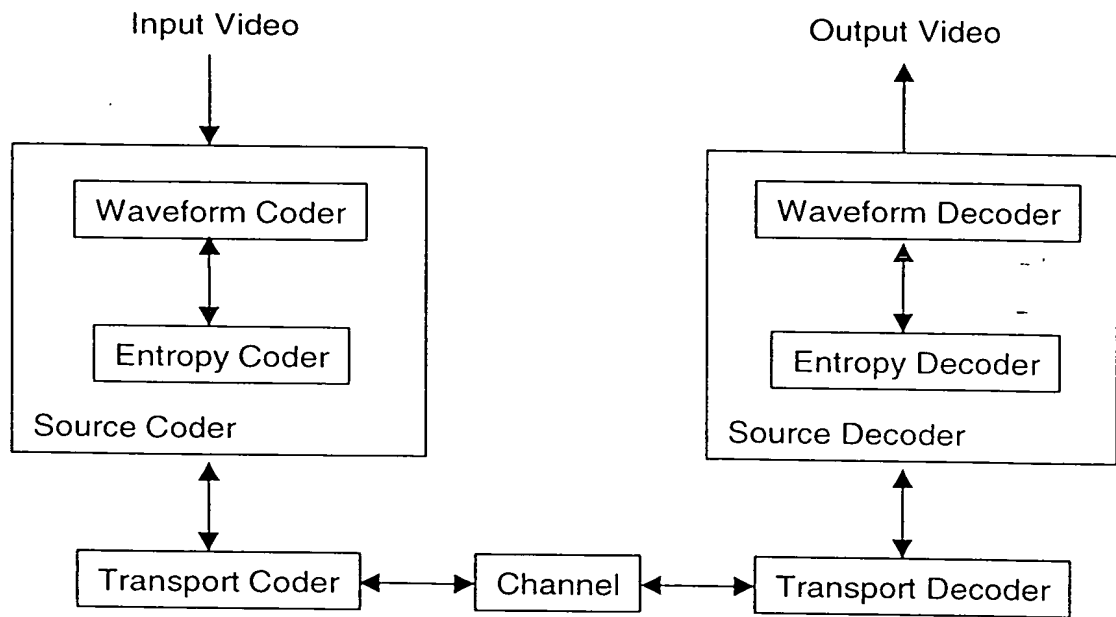


Fig. 1

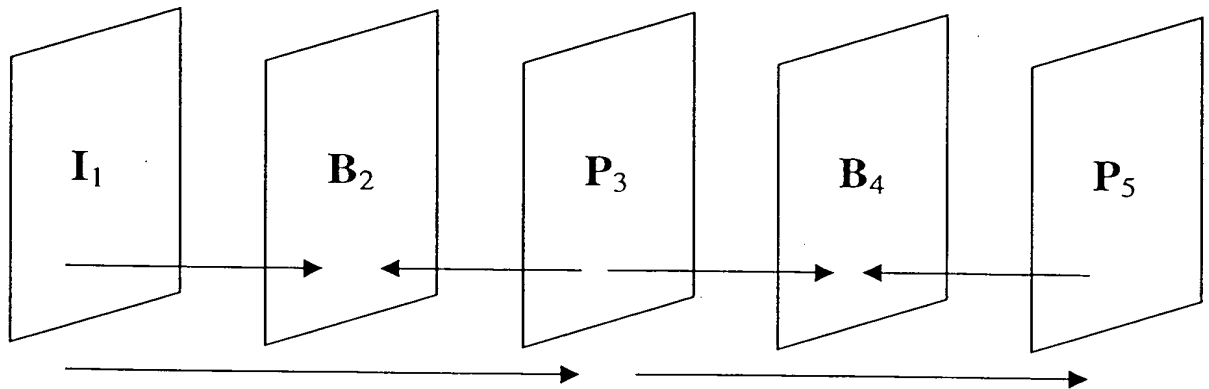


Fig. 2

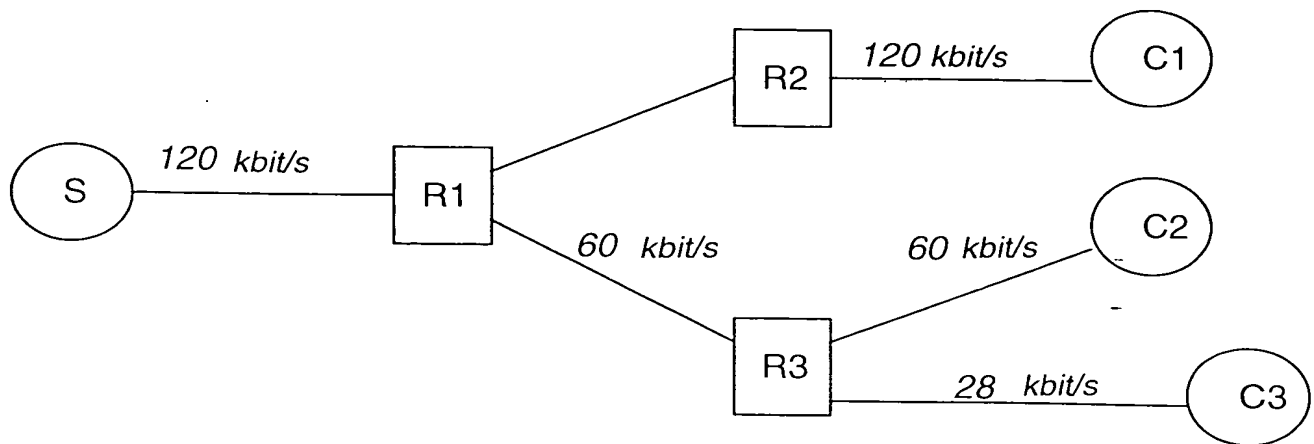


Fig. 3

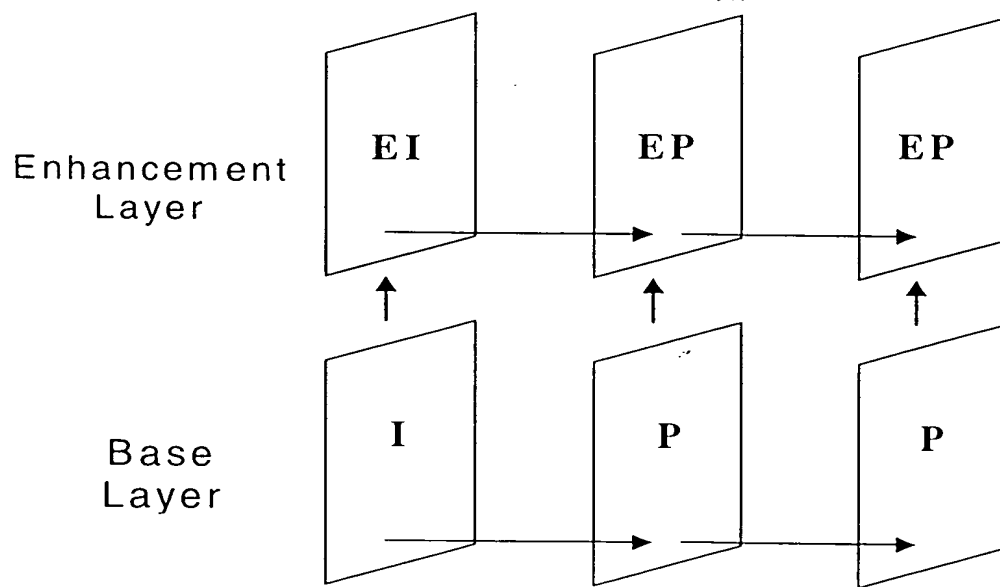


Fig. 4

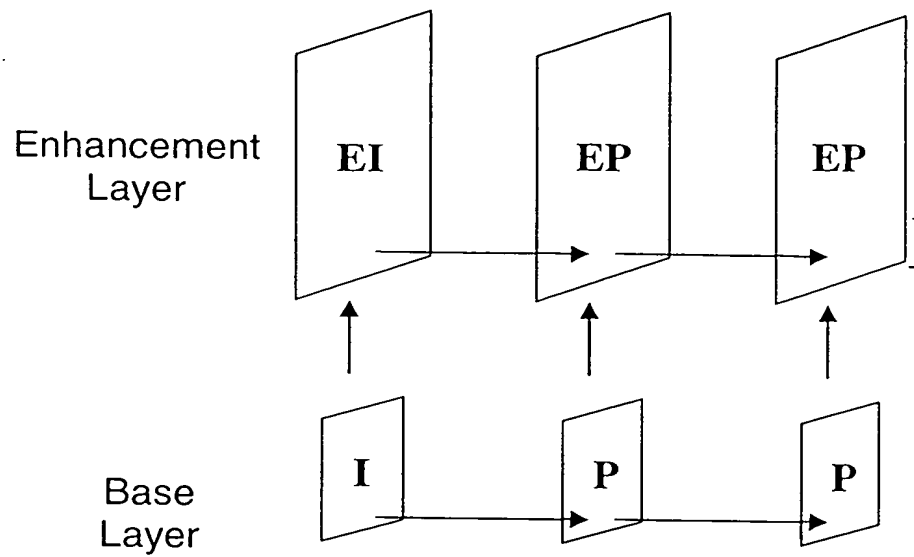


Fig. 5

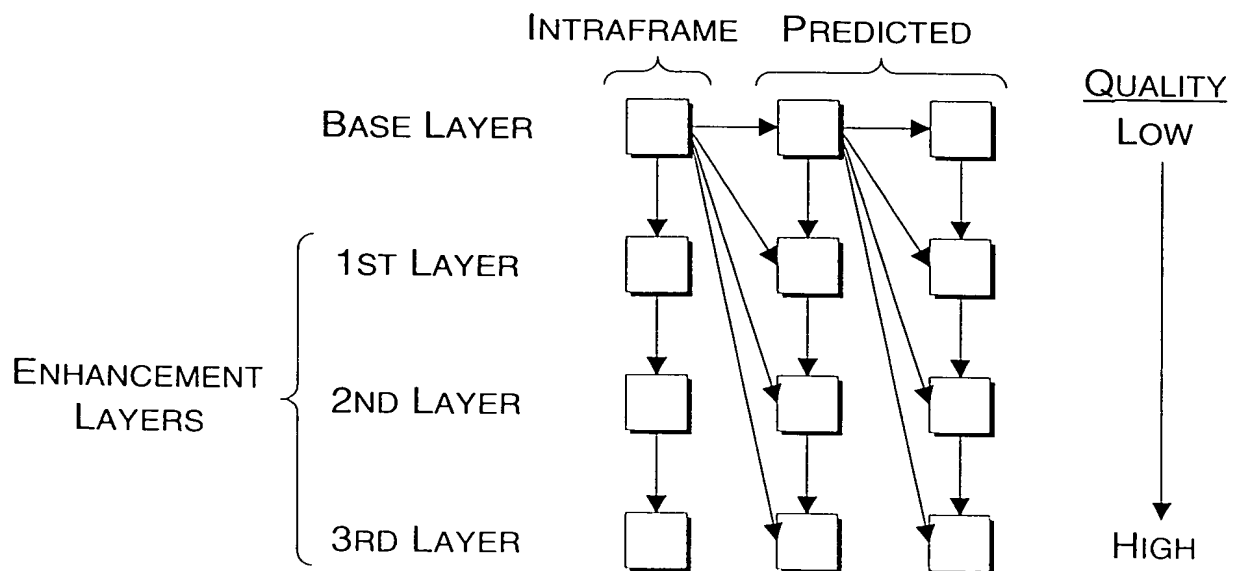


Fig. 6

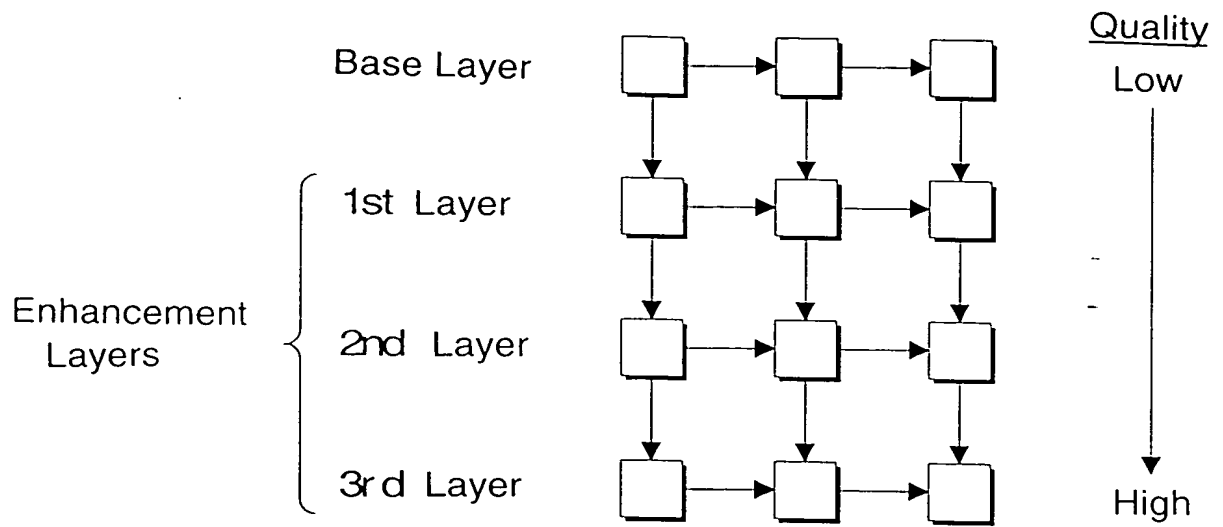


Fig. 7

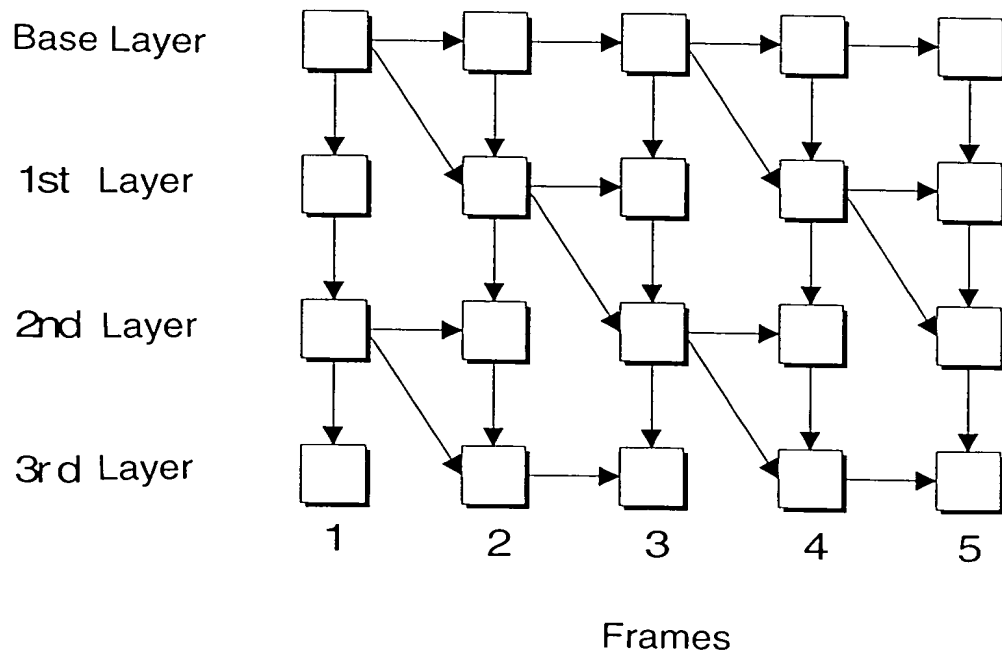


Fig. 8

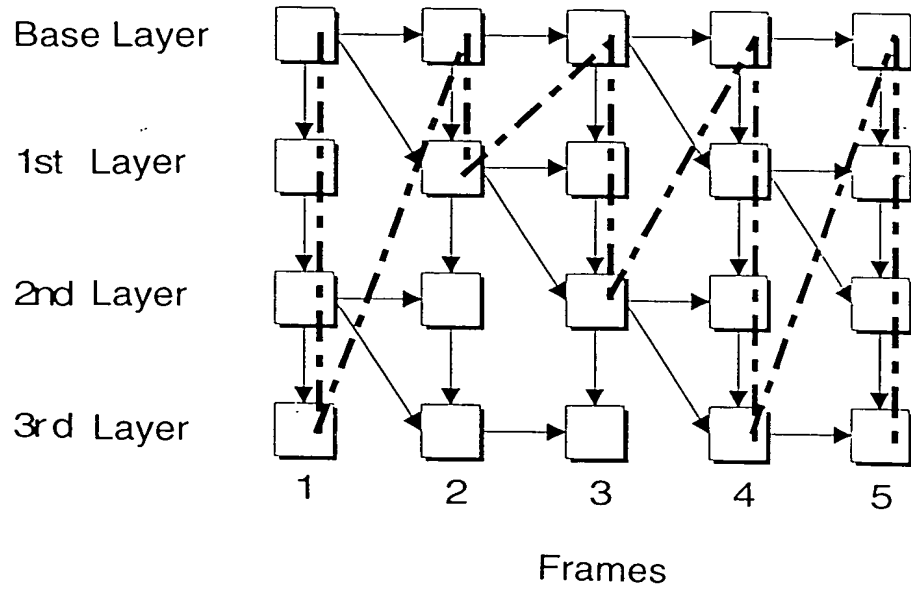


Fig. 9

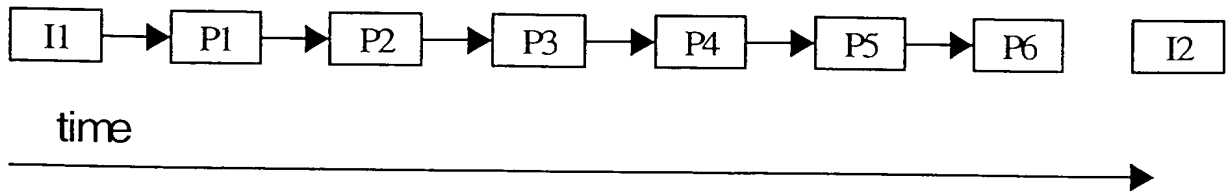


Fig. 10

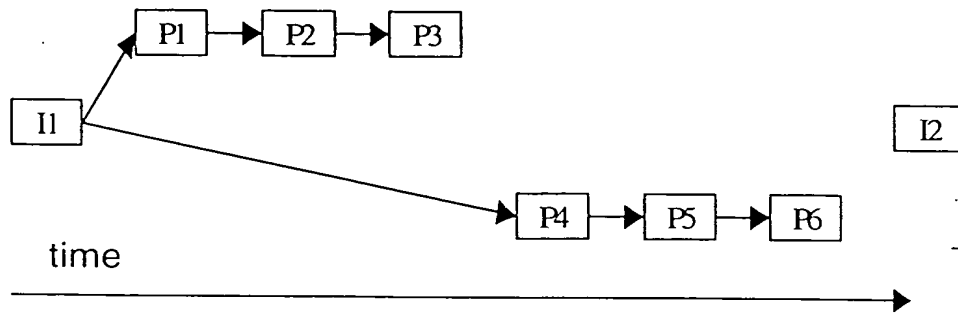


Fig. 11

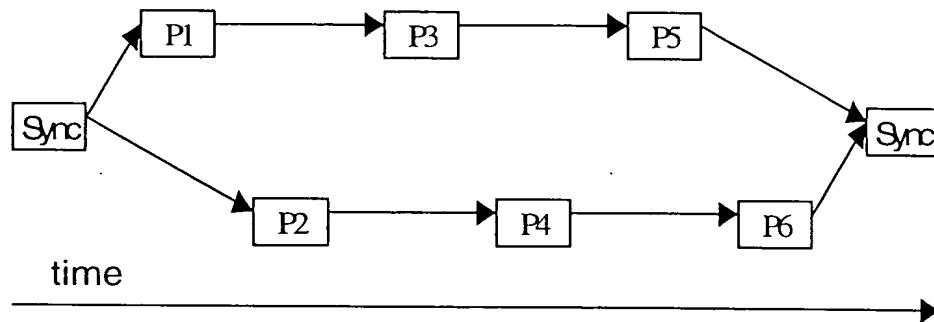


Fig. 12

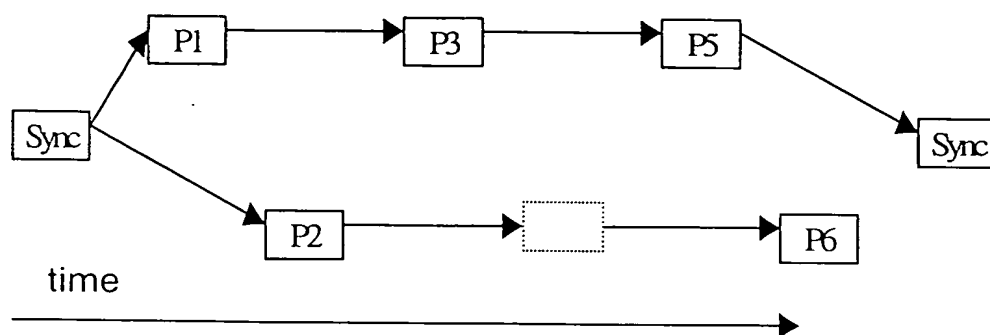


Fig. 13

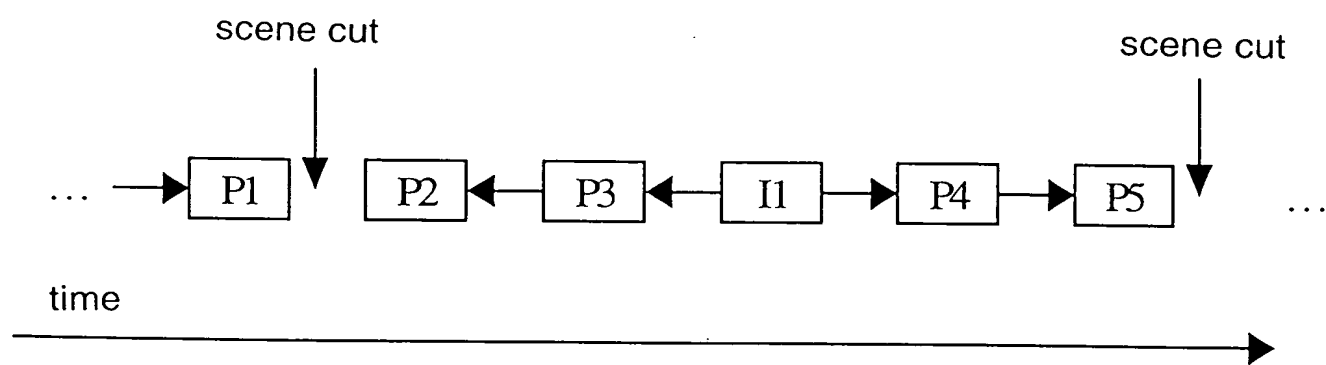


Fig. 14

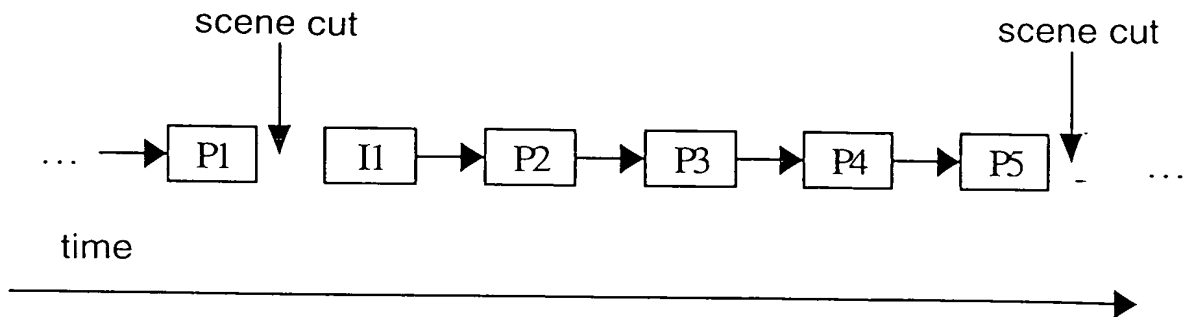


Fig. 15

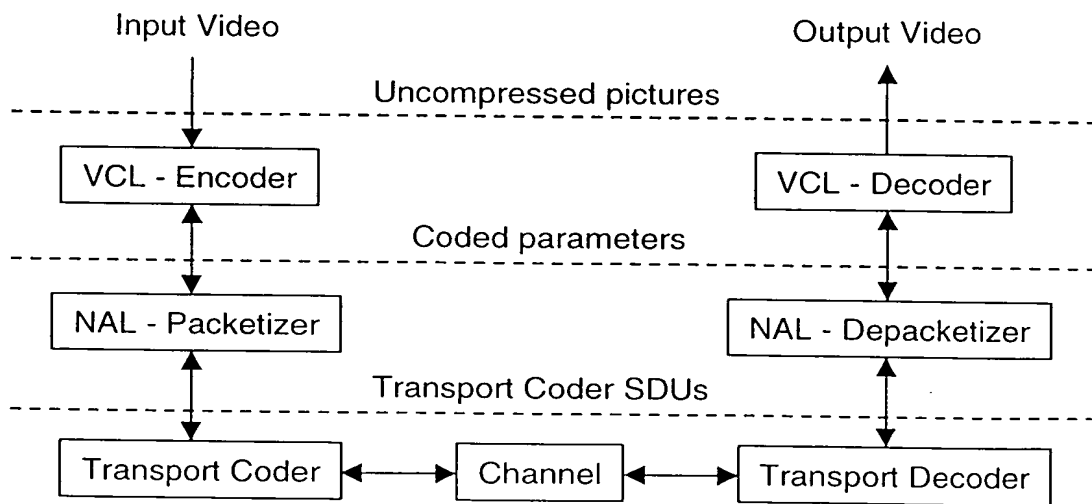


Fig. 16



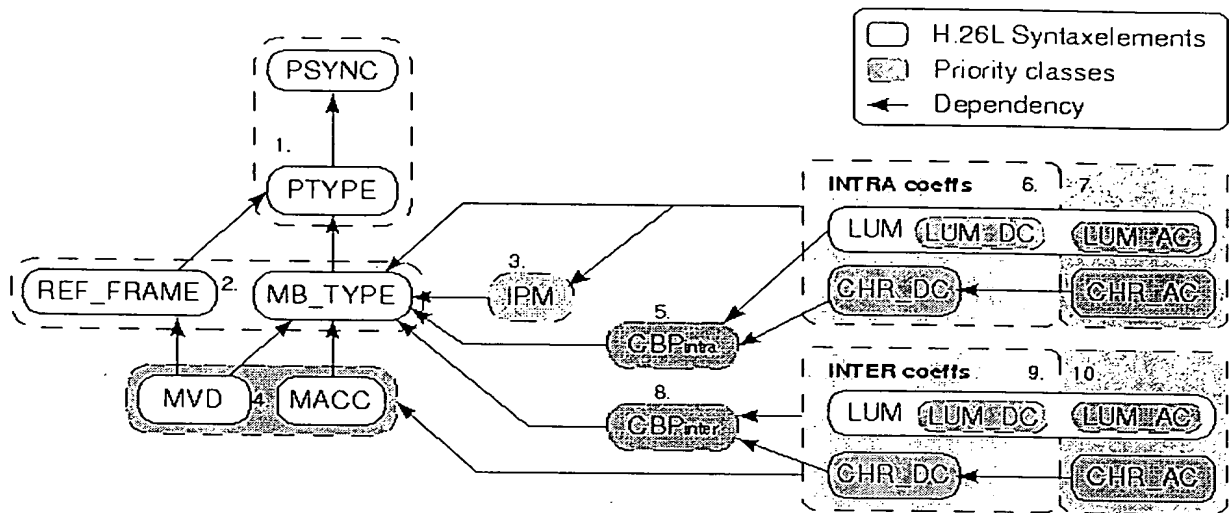


Fig. 17

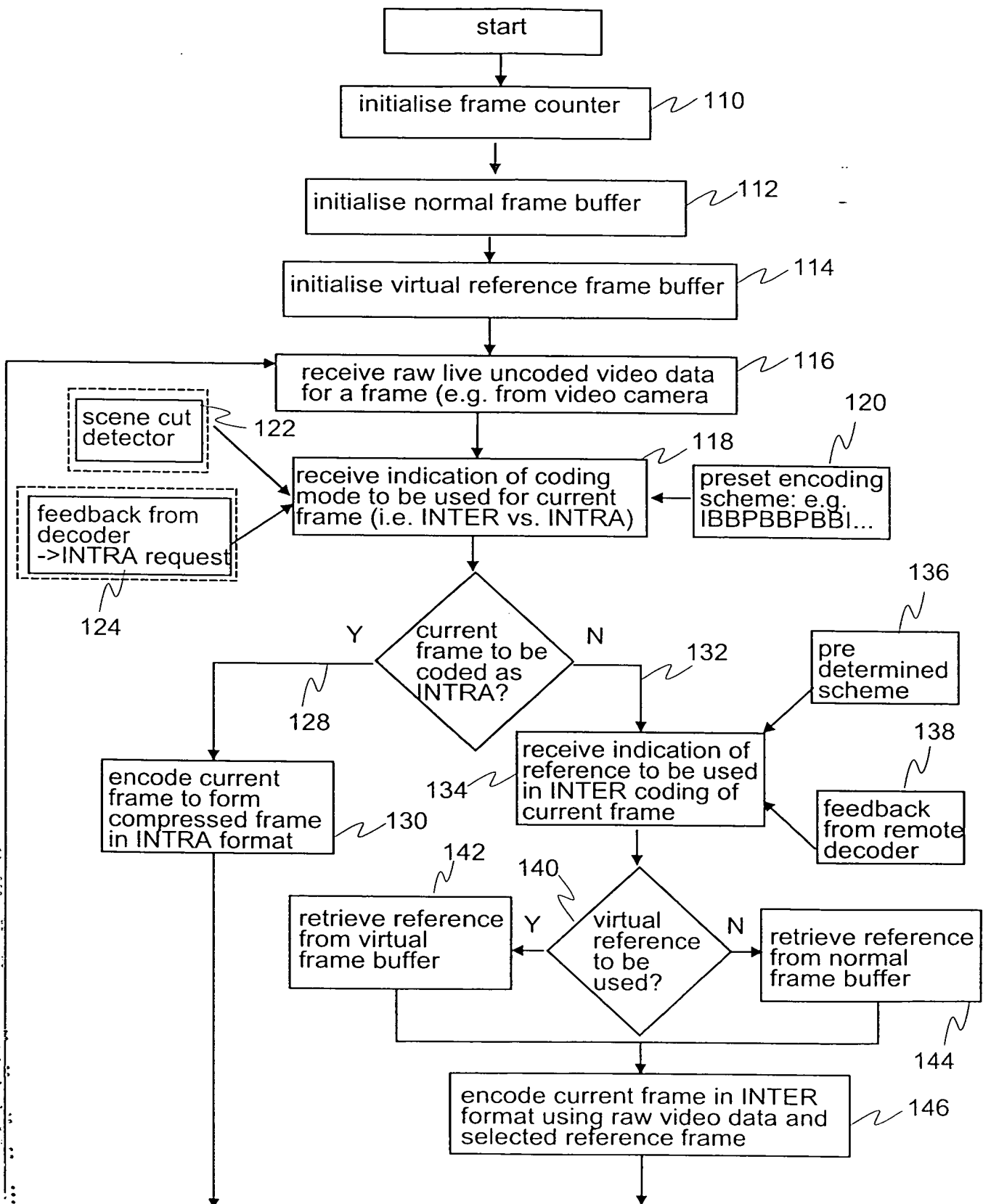


Fig. 18a

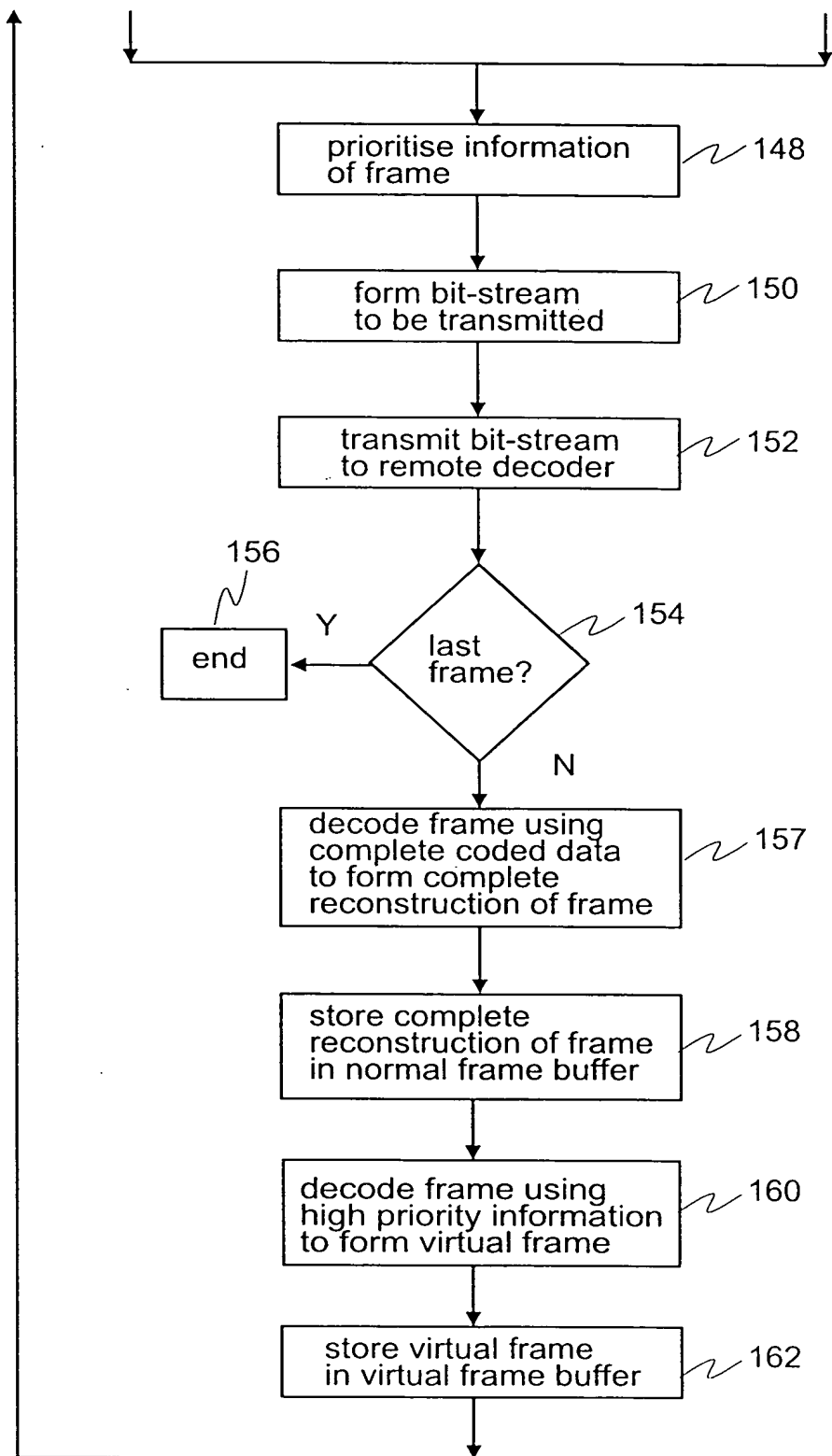


Fig. 18b

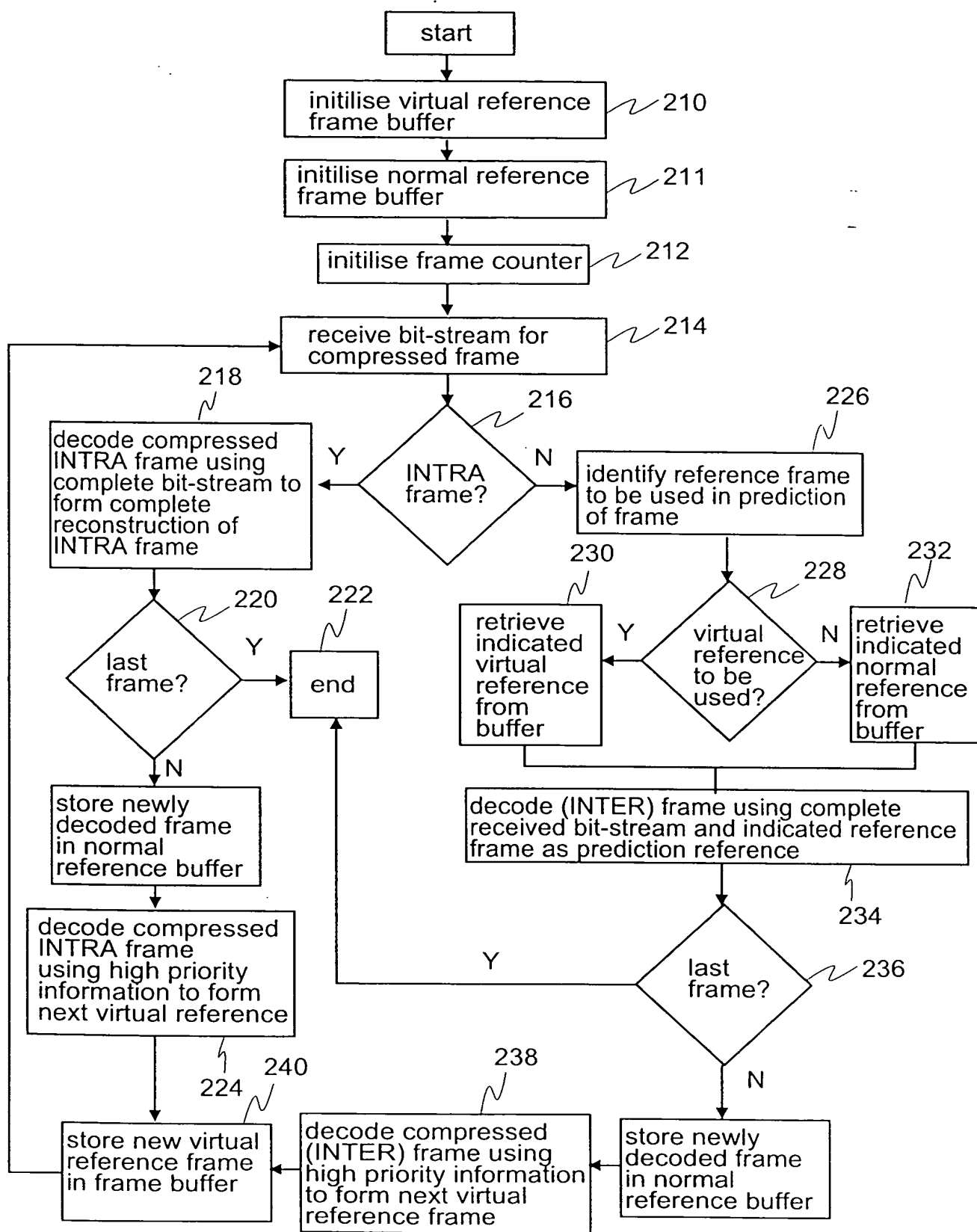


Fig. 19

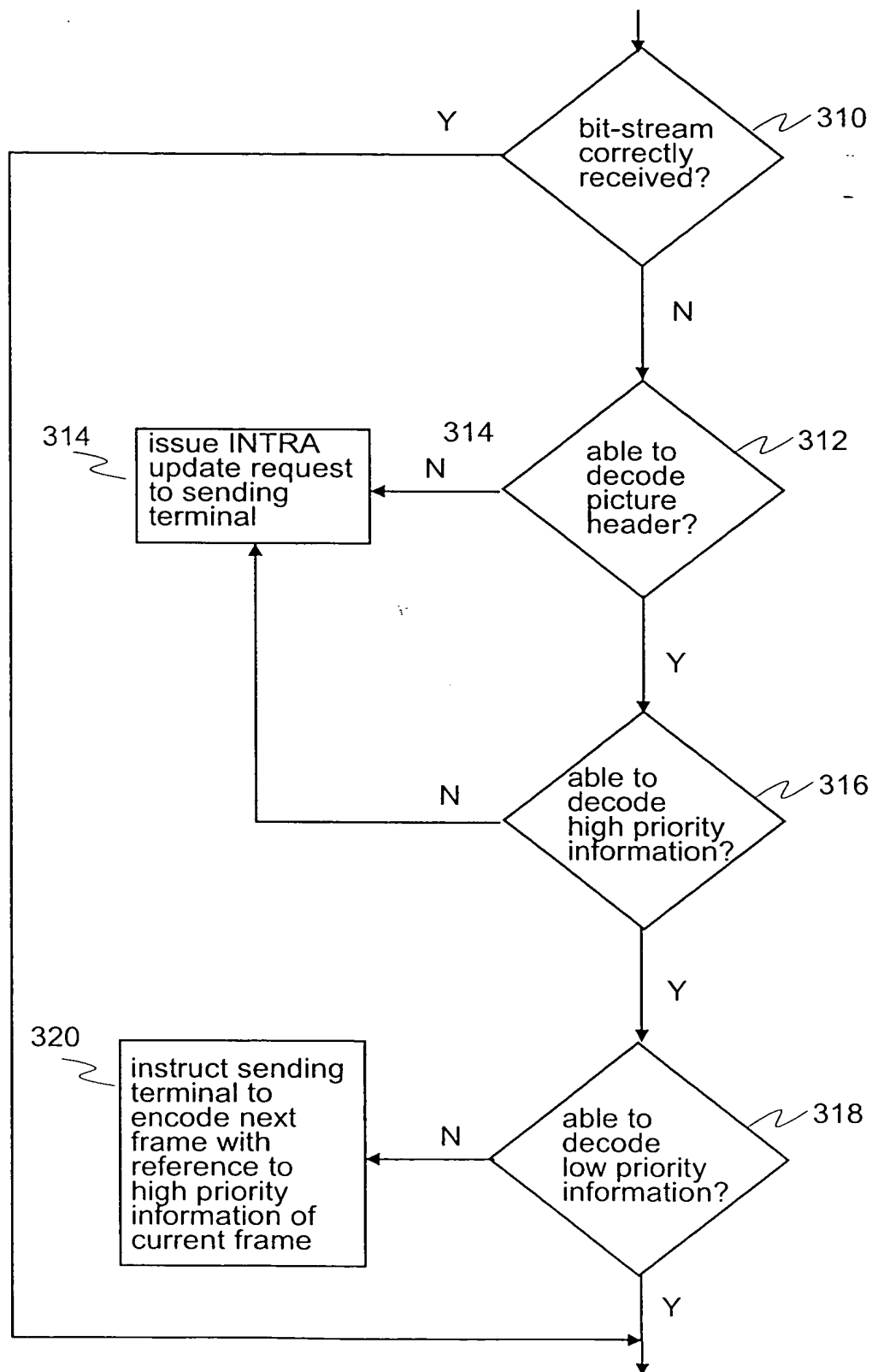


Fig. 20

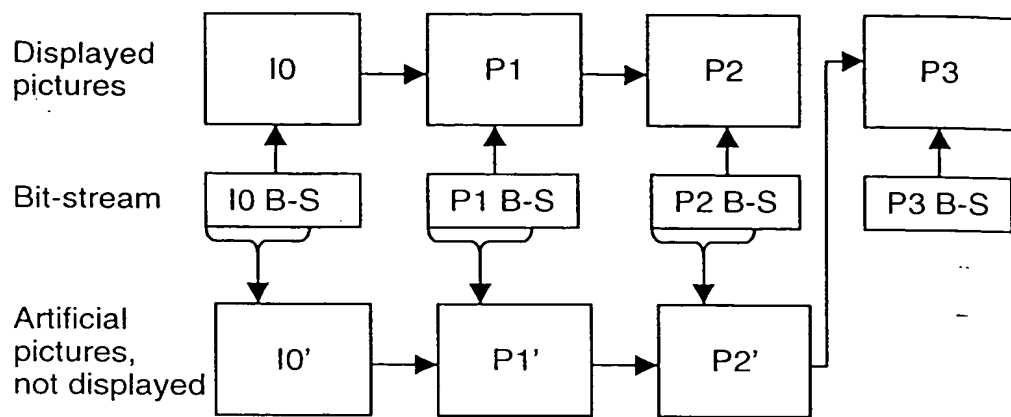


Fig. 21

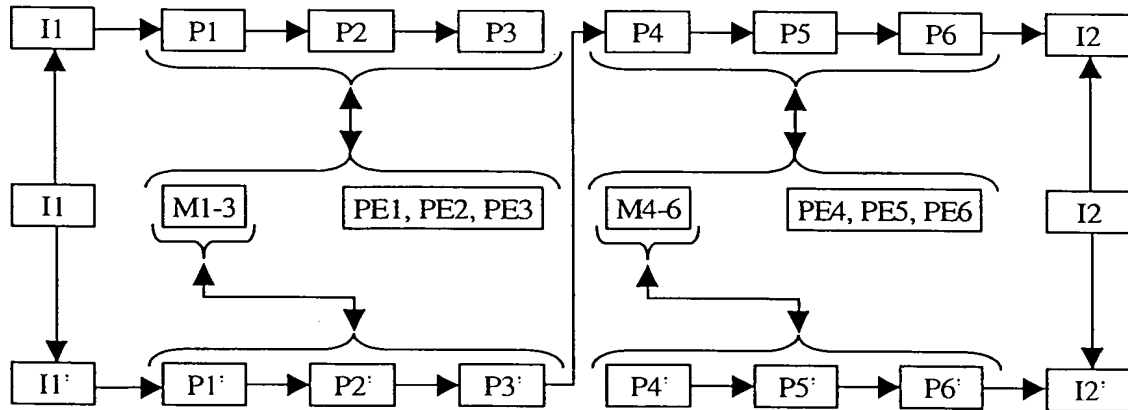


Fig. 22

400

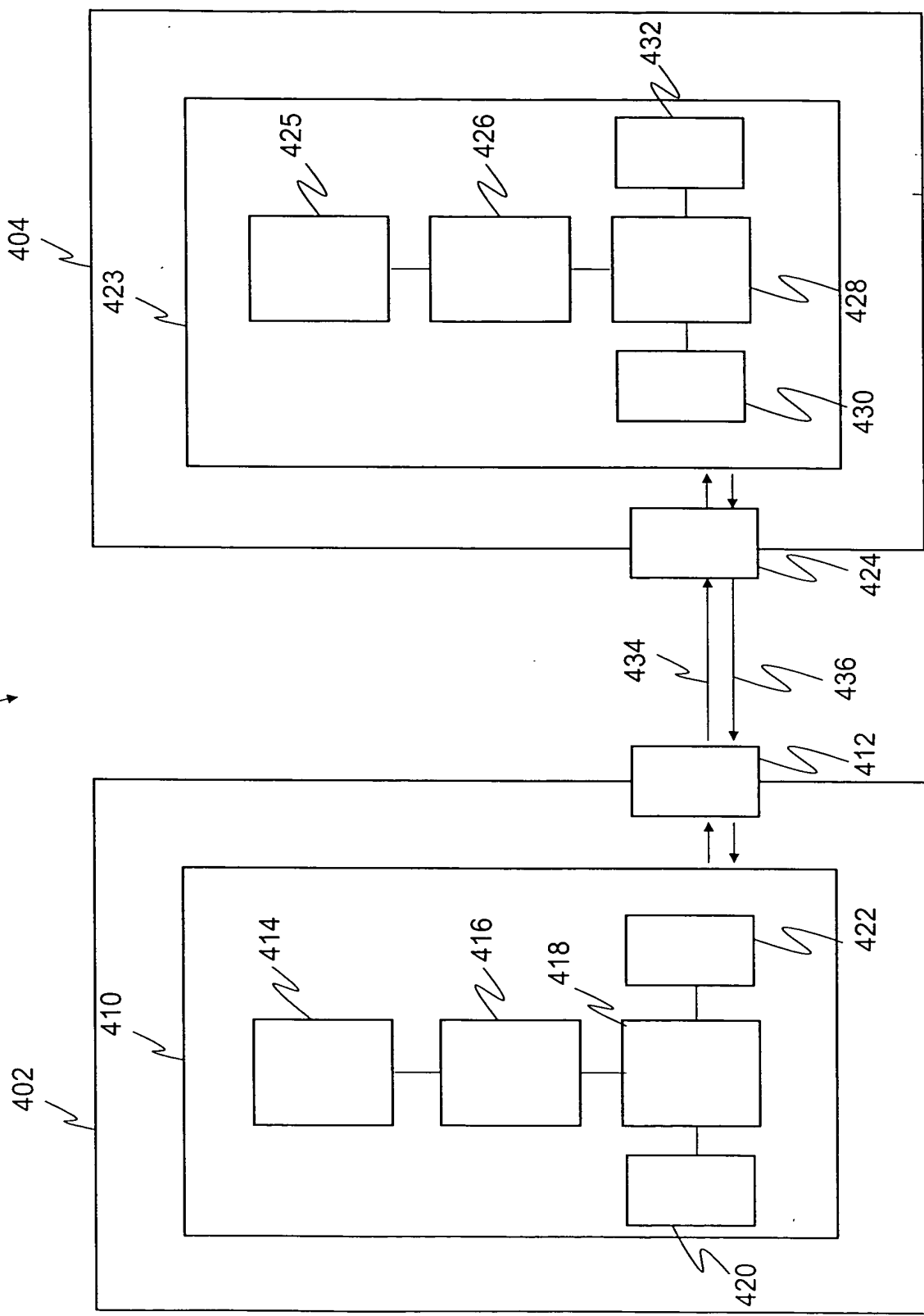


Fig. 23

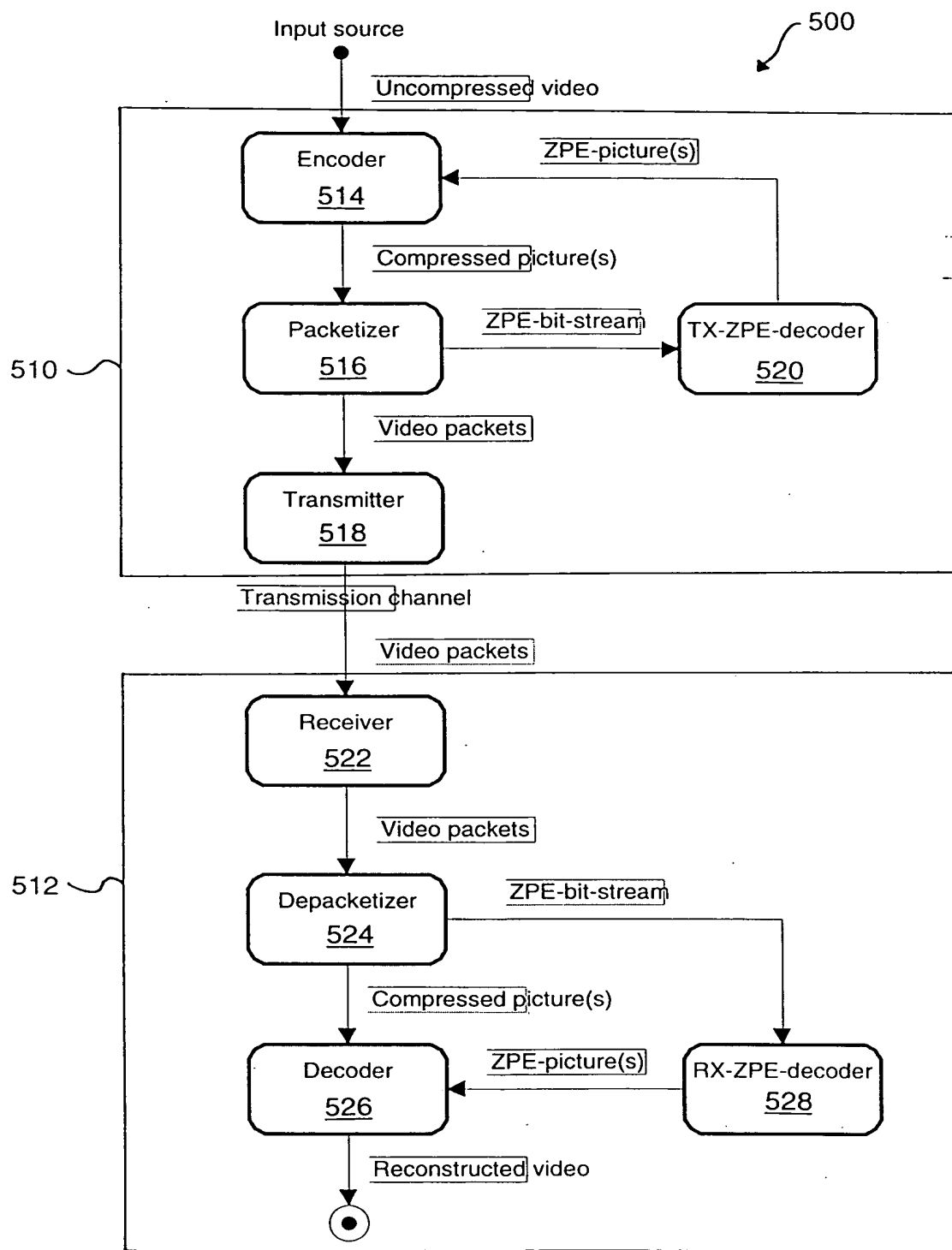


Fig. 24